Bezugssysteme im Leistungsverhalten

Eine experimentelle Studie
zur verhaltenssteuernden Funktion
von Bezugssystemen

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der
Philosophischen Fakultät
der
Westfälischen Wilhelms-Universität
zu
Münster (Westf.)

vorgelegt von

Rainer Brackhane
aus Herford (Westf.)

1976

Bezugssysteme im Leistungsverhalten

Eine experimentelle Studie
zur verhaltenssteuernden Funktion
von Bezugssystemen

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der
Philosophischen Fakultät
der
Westfälischen Wilhelms-Universität
zu
Münster (Westf.)

vorgelegt von
Rainer Brackhane
aus Herford (Westf.)

1976

UNIV.
LEIDEN
BIBL.

Tag der mündlichen Prüfung: 16.7.76

Dekan: Prof. Dr. L. Budde

Referent: Prof. Dr. W. Witte

Korreferent: Prof. Dr. W. Metzger

Bezugssysteme und Leistungsverhalten

Eine experimentelle Studie zur verhaltenssteuernden Funktion von Bezugssystemen

1. Einleitung in den Problembereich 1
1.1. Bezugssysteme und ihre Erforschung 1
1.2. Auseinandersetzung mit Gütemaßstäben 4

2. Aufgabenstellung 7
2.1. Versuchsfragen 7
2.2. Hypothesen 8

3. Zur Kasuistik der verhaltenssteuernden Funktion von Bezugssystemen 9
3.1. Versuchsplanung 9
3.2. Versuchsdurchführung 10
3.3. Ergebnisse 13
3.3.1. Zur Versuchsanordnung und zu allgemeinen Bezugssystem-Charakteristika 13
3.3.2. Zur Auseinandersetzung mit Erfolg und Mißerfolg 20
3.4. Zusammenfassung 26

4. Struktur und Dynamik individueller Leistungsbeurteilungen am Beispiel von Reaktionszeitversuchen 29
4.1. Versuchsplanung 29
4.2. Versuchsdurchführung 30
4.3. Ergebnisse 31
4.4. Zusammenfassung 38

5. Weitergehende Analyse von Struktur und Dynamik der individuellen Leistungseinschätzungen 39
5.1. Versuchsplanung 39
5.2. Versuchsdurchführung 42
5.3. Ergebnisse 43
5.3.1. Zum allgemeinen Versuchs- und Beurteilungsverhalten 43
5.3.2. Vorbemerkungen zur Auswertung der Daten 46
5.3.3. Beurteilung der eigenen Leistung nach verschiedenen Anzahlen von Kategorien 51
5.3.4. Zur Lage der individuellen System-Nullpunkte 75
5.3.5. Anspruchsniveau-Setzungen und Anspruchsniveau-Verschiebungen 80
5.4. Zusammenfassung 83

6. Schlußbetrachtung 87
6.1. Zusammenfassende Erörterung 87
6.2. Kritische Rückschau 97
6.3. Ausblick 102

Literatur-Verzeichnis 106

Danksagung 109

Lebenslauf und Bildungsgang 110

Anhang 111

1. Einleitung in den Problembereich

1.1. Bezugssysteme und ihre Erforschung

Für alles, was uns auf den verschiedenen Ebenen unseres Erlebens je begegnet ist, besitzen wir mehr oder minder fest ausgeprägte "Bezugssysteme". Sie befähigen uns dazu, die weitaus meisten neuen Erlebnisse schnell und zuverlässig einzuordnen, was wiederum Voraussetzung ist für das Zurechtfinden in unserer Welt - wir können uns bei der Begegnung mit wirklich Neuem diesem mit unserer ganzen Aufmerksamkeit zuwenden.

Unser Umgang mit unseren jeweiligen Bezugssystemen zeigt sich etwa darin, daß wir unwillkürlich absolut urteilen, ja den Dingen selbst sogar absolute Eigenschaften zusprechen: ein Baum ist hoch, ein Weg weit, ein Bild finden wir schön, ein Redner spricht gut.
Derartige absolute Eigenschaften (oder sie kennzeichnende Urteile) werden also nicht als "auf irgendetwas anderes bezogen" erlebt oder "aus dem Eindruck selbst" verständlich werden, sondern sind nur zu verstehen aus der funktionalen Abhängigkeit von einem System, das sich wiederum "ausschließlich durch die aus diesem Bezug vorhergesagten Eigenschaften bestätigt" (WITTE 1960a). Diese "Bezugssysteme" (vgl. u.a. WITTE 1974) bedingen normalerweise eine starke Tendenz zur Verabsolutierung von Eigenschaften und sie bezeichnenden Urteilen, die sich auch etwa gegen ausdrücklich anderslautende Instruktionen durchsetzen kann. So wurde dieses Phänomen denn auch ursprünglich - eher beiläufig - als eigentlich nicht-anweisungsgemäßes Verhalten von Versuchspersonen entdeckt bzw. wissenschaftlich aktualisiert (MARTIN & MÜLLER 1899).

Seitdem man danach auf die Rolle der Bezugssysteme im täglichen Leben aufmerksam geworden war, wurden sie in wachsendem Ausmaße Gegenstand des Interesses. Seit der ersten Kartographie dieses Phänomenbereiches durch WEVER & ZENER (1928) und einer ersten systematischen Darstellung durch METZGER (1940, 5.Aufl. 1975) ist diesem Thema eine Mehrzahl von Darstellungen und Versuchsreihen gewidmet worden. Genannt seien hier vor allem die Namen HELSON, STEVENS und WITTE. Es wurden einige ganz voneinander verschiedene Forschungsansätze unternommen - teilweise stand die Entstehung von Bezugssystemen im Vordergrund (Adaptationsprozesse), teilweise ging es primär um die Struktureigenschaften von bereits ausgebildeten (mnestisch stabilisierten) Bezugssystemen; als Grundlage der Absoluturteile wurden sensorische oder semantische Faktoren oder das Phänomen der Ähnlichkeit angenommen - eine Zusammenschau hat sich bisher noch als nicht möglich erwiesen (vgl. SITMALIDIS 1974), lediglich innerhalb der einzelnen Ausrichtungen kommt es zu immer besseren Absicherungen (vgl. FORNER 1974).

Wir wollen an dieser Stelle nicht einen neuerlichen Ansatz zur Integration unternehmen, sondern vielmehr versuchen, auf dem Boden der Erforschung der mnestisch stabilisierten Bezugssysteme, wie sie WITTE und seine Mitarbeiter seit einer Reihe von Jahren verfolgt haben, den Geltungsbereich der dort gefundenen Gesetzmäßigkeiten auszuweiten - und dies in zweierlei Weise:
Die bisherigen Untersuchungen innerhalb dieses Ansatzes wurden fast durchweg an Gegenstands-Mannigfaltigkeiten vorgenommen, die über physikalische Metriken (z.B. cm, g, Hz) einzuordnen sind (Bleistifte, Striche, Armbanduhren, Gewichte, Töne u.a.). Auf der Grundlage der dabei gewonnenen Ergebnisse formulierte WITTE (1960a) die folgenden Gesetzmässigkeiten:

1. Die Systeme sind ihrem absoluten Umfang ebenso wie die Bereiche ihrem relativen Umfang nach normalverteilt.
2. Die Systemenden sind normalverteilt.
3. Jeder relative Bereichsumfang ist unabhängig vom absoluten Systemumfang.
4. Systemanfang und -ende sind unabhängig voneinander.
5. Systemanfang und -ende sind unabhängig von den relativen Bereichsgrößen.
6. Die intraindividuelle Streuung ist überaus signifikant kleiner als die interindividuelle Streuung.

(nach FORNER 1974).

Einzig KELLNER (1966, vgl. WITTE 1971) und SKROBLIN (1972) gingen hier andere Wege, indem sie Bezugssystem-Strukturen auf nicht-sensorischen Gebieten untersuchten. Gleichsam auf ihren Spuren wollen wir in dieser Arbeit sehen, ob sich die bisher gefundenen Verhältnisse reproduzieren lassen auf einen Bereich, bei dem man keine physikalische Metrik angeben kann, nämlich dem des allgemeinen Verhaltens (vgl. SIEVERS 1970).

Wir gehen damit gleichzeitig eine weitere Funktion der Bezugssysteme für unser Erleben und Handeln an:

Bisher war überwiegend - wie dies auch in den oben genannten Beispielen zum Ausdruck kommt - von ihrer orientierenden Funktion die Rede, die verständlich wird aus dem Bemühen des Individuums heraus, auch "im geistigen Gelände" die Orientierung zu behalten, "um den Stellenwert des Ortes, an dem man sich befindet", zu wissen (vgl. WITTE 1960a). Darüber hinaus hat etwa SKROBLIN (1972) herausgestellt, daß bei aller verschiedenen Erlebensweise die Übereinstimmung im Kollektiv so groß ist, daß die Bezugssysteme über den Selbstorientierungswert hinaus auch Kommunikationswert haben: Die Mitteilung eines Sprechers, er habe einen tiefen Ton gehört, wird bei seinem Gesprächspartner kaum die Vorstellung vom dreigestrichenen C auslösen.

Bereits bei WITTE (1960a) wird als dritte Funktion der Bezugssysteme die verhaltenssteuernde deutlich, wenn er "eine absolute Eigenschaft" neben dem "sie festhaltenden Urteil" auch mit dem "ihr entsprechenden Verhalten" parallel setzt. Alle drei werden nur verständlich aus der oben bereits angesprochenen "funktionalen Abhängigkeit vom gesamten System". Dies deutet auch SKROBLIN an, wenn sie für die Ausbildung von Bezugssystemen - und so für ihre Struktur - die Implikation von Motivation fordert, Beweggrund für eine ganz bestimmte Art und Weise von Verhalten.

1.2. Auseinandersetzung mit Gütemaßstäben

Eines der lohnendsten Felder bei der angezielten Untersuchung der verhaltenssteuernden Funktion von Bezugssystemen dürfte das Gebiet der Leistungsmotivation sein, der "Auseinandersetzung mit Gütemaßstäben" (McCLELLAND u.a. 1953). Lohnend vor allem deshalb, weil der Leistungsbereich ein wesentlicher Aspekt menschlichen Verhaltens ist, der zudem definitorisch wie experimentell einigermaßen gut in den Griff zu bekommen ist. Folgt man HECKHAUSEN (1963a), dann muß man sich sogar wundern, daß die Ergebnisse der Bezugssystemforschung nicht längst eingehender auf die Inhalte der Motivationspsychologie angewendet worden sind, speziell der Leistungsmotivation. Die einschlägigen Begriffe sind imgrunde vorhanden, absolute Urteile auch hier eine Selbstverständlichkeit: in unseren einzelnen Tätigkeiten werden meist - mindestens persönliche - Gütemaßstäbe für verbindlich gehalten, bei ihrer Ausführung wird allgemein präzise und sicher zwischen Gelingen und Mißlingen, zwischen Erfolg und Mißerfolg unterschieden, es ist in allen relevanten Fällen ein individuell bevorzugter Gütegrad (Anspruchsniveau) vorhanden, an dem alles gemessen wird.

Wie groß die Nähe zwischen den Inhalten der Bezugssystem-

Forschung und ihren Äquivalenten in der Motivationspsychologie tatsächlich ist, mag aus dem folgenden kurzen Abriß noch deutlicher werden, vor allem an der Terminologie. Wir lehnen uns dabei im wesentlichen an HECKHAUSEN an, speziell an seine Motivationsthesen (1963a).

"Motivationen beruhen auf hochorganisierten kognitiven Systemen von Person-Umwelt-Bezügen." (These 1). "Diese Systeme bauen sich vor allem aus relativ überdauernden Erfahrungsniederschlägen ... auf, bleiben jedoch mehr oder weniger erfahrungsabhängig, d.h. wandlungsfähig." (2). "Jedes System stellt mit seinem verallgemeinerten Grundgerüst ein Bezugssystem für den jeweilig aktuellen Person-Umwelt-Bezug einer gegebenen Inhaltsklasse dar." (z.B. Leistungsverhalten) "Es legt insbesondere die Norm fest, wie ein Umwelt-Bezug beschaffen sein muß, um sich als ein befriedigender vor anderen auszuzeichnen." (4). "Die Norm eines befriedigenden Person-Umwelt-Bezuges - oder kurz 'Soll-Lage' - wird durch verschiedene Faktoren festgelegt" (5), je nach Gewicht dieser Faktoren können die Soll-Lagen verschiedener Natur sein (6):

> "etwa durchschnittsstatistische Normal-Lagen nach Art sich wandelnder Adaptationsniveaus (wenn Erfahrungsniederschläge ausschlaggebend sind); oder vorgegebene bzw. relativ fixierte Optimal-Lagen (wenn konstitutionelle Festgelegtheiten des Organismus ausschlaggebend sind); oder Ideal-Lagen, die weniger ein fixiertes Niveau als eine Richtungstendenz, ein Gefälle auf einen ausgezeichneten Prägnanzzustand hin darstellen (wenn sachimmanente Prägnanzstrukturen oder tradierte Normvorstellungen ausschlaggebend sind); oder Pendel-Lagen mit einem alternierenden Zyklus von Spannung und Lösung (wenn ein Aktivierungszirkel auslösendes Anregungspotential ausschlaggebend ist)."

HECKHAUSEN selbst führt hierzu weiter aus, bei der Klassifikation von Soll-Lagen gehe es um das allgemeinpsychologische Problem der Bezugssysteme und ihrer Genese, das für die Motivationspsychologie nicht weniger grundlegend sei als für die Wahrnehmungspsychologie. Gerade leistungsmotiviertes Verhalten liefere genügend Phänomene von Soll-Lagen, die nur

als ein eigenartiges Produkt von Normal- und Ideal-Lagen zu begreifen seien.

Wie diese Soll-Lagen durch ihre Diskrepanzen zu den Ist-Lagen (d.h. den "gegenwärtig bestehenden oder ... als künftige im Erleben vorweggenommenen Person-Umwelt-Bezüge(n)" - These 7) in der individuellen Auseinandersetzung mit der eigenen Leistung wirksam werden bzw. in welcher Abfolge sich verschiedene Soll-Lagen manifestieren, soll uns hier im folgenden beschäftigen.

Abweichend von HECKHAUSEN schränken wir dabei den Begriff der Ist-Lage ein, weil uns dies terminologisch klarer erscheint. Wir folgen dabei FUCHS (1963), indem wir jede Bezugslage als Sollage bezeichnen. "Der Terminus Istlage bleibt reserviert für das gegenwärtig Bestehende, das in Bezug auf eine Sollage organisiert ist." HECKHAUSEN bezeichnet dagegen als Sollagen nur die Lagennormen, die generalisierten Normen eines befriedigenden Umweltbezuges, und stellt dem gegenwärtige und künftige Istlagen gegenüber, also auch die Ziele eines Umweltbezuges.

2. Aufgabenstellung

2.1. Versuchsfragen

Die angesprochene Auseinandersetzung mit der eigenen Leistung spielt sich ab im Rahmen der Leistungsmotivation, die HECKHAUSEN (1965) definiert "als das Bestreben, die eigene Tüchtigkeit in allen jenen Tätigkeiten zu steigern oder möglichst hoch zu halten, in denen man einen Gütemaßstab für verbindlich hält und deren Ausführung deshalb gelingen oder mißlingen kann. ... Relevante Ereignisse eines leistungsthematischen Person-Umwelt-Bezuges - und damit zeitliche Endzustände, auf die sich Leistungsmotivation bezieht - sind letztlich immer Erfolg und Mißerfolg." Und er fährt fort: "Wie differenziert auch ein Gütemaßstab sein mag, er zerfällt in einen Teil, der - wie ausgeprägt auch immer - Erfolg, und in einen anderen, der Mißerfolg bedeutet. ... In Abhängigkeit vom erzielten Leistungsstand wandert ... der konkretisierte Nullpunkt des Gütemaßstabes."
Die hierbei zugrunde liegenden Strukturen gilt es im folgenden zu untersuchen, wie überhaupt die Einordnung dieser Arbeit schon aus dem im vorigen Kapitel Gesagten deutlich geworden sein mag:
In welcher Form und inwieweit führen Leistungsveränderungen zu einer Wanderung des Nullpunktes des relevanten Gütemaßstabes, zu einer Bezugssystem-Änderung also?
Welche Strukturen von Beurteilungen der eigenen Leistung lassen sich nachweisen, welche Strukturen auch von Erwartungssystemen hinsichtlich der eigenen Leistung, des eigenen Verhaltens?
Welcher Art sind die zugrunde liegenden Normen? Stehen die für die Aufgabenbeschreibung einschlägigen Kategorien leicht/schwierig in einem greifbaren Zusammenhang mit den Begriffen

gut/schlecht, die zur Beurteilung der eigenen Leistung herangezogen werden?
Vor allen Dingen aus diesen letzten Sätzen mag deutlich werden, welche Rolle die von uns unwillkürlich immer wieder verwendeten Bezugssysteme auch und gerade für das Verständnis unseres Verhaltens haben, wie sehr gerade das Erleben von Erfolg und Mißerfolg auf unser Verhalten steuernd wirkt - gleichgültig wohl, ob dem eine vorwiegend kognitive oder eine eher emotionale Orientierung zugrunde liegt.

2.2. Hypothesen

Wir vermuten, daß wir verschiedene Arten von Bezugssystemen, von Gütemaßstäben für Erleben und Beurteilung des eigenen Leistungsverhaltens intra- und interindividueller Form nachweisen können, die ganz verschiedener Genese, aber ähnlicher Struktur sein mögen.

Wir gehen mit HECKHAUSEN davon aus, daß (im konkreten Fall der Leistungsmotivation) bei einem Individuum sich aufgrund einer vorgegebenen Aufgabe die unterschiedlichen Arten von Soll-Lagen bzw. von Istlagen-Sollagen-Diskrepanzen derartig in spezifischen Anspruchsniveausetzungen manifestieren, daß sich in den Zielsetzungen und den ihnen entsprechenden Handlungen die Diskrepanz zur "befriedigenden Norm" ablesen läßt.

Wir vermuten schließlich, daß sich auch für die individuelle Beurteilung des eigenen Leistungsverhaltens und damit wohl auch für das Leistungsverhalten selbst auf der Folie des Erlebens von Erfolg oder Mißerfolg die Gültigkeit von WITTEs Bezugssystem-Theoremen nachweisen läßt, die bisher in erster Linie auf verschiedenen sensorischen Gebieten erhärtet worden ist.

3. Zur Kasuistik der verhaltenssteuernden Funktion von Bezugssystemen

3.1. Versuchsplanung

Bei den Überlegungen über einen geeigneten Versuchsplan zur Beantwortung der gestellten Fragen kristallisierte sich bald heraus, daß es empfehlenswert sei, angesichts des Erkundungscharakters dieser Untersuchung zwei Schwerpunkte zu setzen:
Erstes Ziel sollte dabei sein, auf dem ausgesuchten Untersuchungsfelde auf dem Wege über eine kasuistische Zusammenstellung von Phänomenen näheren Einblick in die Variabilität des Feldes und damit auch in die Vielfältigkeit der Verhaltensweisen unserer Versuchspersonen (Vpn) zu erhalten.
Dies stellte sich als notwendig dar, weil erst auf dieser Grundlage eigentlich nähere Aussagen zur Art möglicher Bezugssysteme und Bezugssystemänderungen würden gemacht werden können. Wir mußten zunächst einmal wissen, welcher Art die zu untersuchenden Bezugssysteme überhaupt waren, um dann auf ihre Struktur und Dynamik eingehen zu können.

Für die damit angesprochene erste Versuchsserie war es also im Sinne dieser Überlegungen notwendig, den Versuchspersonen im Versuch möglichst viel Spielraum einzuräumen, die Untersuchung möglichst qualitativ anzulegen, um möglichst viel von dem Spektrum der für die Soll-Lagen-Bestimmung maßgeblichen Determinanten durchschaubar zu machen.
Ein besonderes Anliegen mußte weiter sein - da ja die Untersuchung auf dem Felde der Leistungsmotivation durchgeführt werden sollte - eine Versuchsanordnung zu finden, die die notwendige Leistungsatmosphäre sozusagen von sich aus mitbrachte, damit das ganze Augenmerk den Erlebens- und Beurteilungsstrukturen der Versuchspersonen gewidmet sein konn-

te. Wir entschieden uns in diesem Sinne für eine erste Versuchsserie mit dem Dart-Spiel:

> Es wird mit Pfeilen auf eine runde Scheibe von 44 cm Durchmesser geworfen, die gleichabständig abwechselnd schwarze und gelbe Ringe mit den aufgedruckten Werten 10, 25, 50, 75 und 100 aufweist.
> Das Zentrum der Scheibe bildet im schwarzen 100-Kreis ein roter Punkt von etwa 1 cm Durchmesser, dem man den Wert 150 zuschreiben kann.
> Zur handelsüblichen Ausfertigung der Scheibe werden sechs Pfeile geliefert, je zwei in blau, rot und gelb.

Der nahezu unwiderstehliche Aufforderungscharakter dieser Wurfscheibe in Verbindung mit ihrer klaren - auch für die Versuchspersonen jederzeit übersehbaren - Struktur konnte nur im Sinne unserer Arbeit sein.

Hinsichtlich der Auswahl unserer Versuchspersonen entschieden wir uns auf dem Hintergrund der oben ausgeführten Überlegungen dafür,

(a) möglichst introspektions- und verbalisierungsfähige Vpn auszuwählen, die bei aller Freude am Spiel auch zur weitergehenden Mitarbeit bereit wären;

(b) die Vpn nach Möglichkeit aus dem persönlichen Bekanntenkreis auszuwählen, um möglichst weitgehend über sie verfügen zu können;

(c) die Versuche dementsprechend - und entsprechend der Spielatmosphäre - in privater Umgebung durchzuführen;

(d) die Versuchsanweisungen möglichst offen zu geben, um das Verhaltensspektrum der Vpn nur soweit einzuschränken, als dies unumgänglich war.

In späteren Versuchsserien werden dann die hier gewonnenen Erkenntnisse abzusichern und näher aufzuschlüsseln sein.

3.2. Versuchsdurchführung

Die Pfeilwurf-Versuche der ersten Serie wurden im Sinne der geschilderten Planung mit Partnern aus dem persönlichen Be-

kanntenkreis - meist Studienkollegen - in privater Umgebung durchgeführt; allen Versuchspersonen war das Dart-Spiel nicht nur bekannt, sondern gut vertraut. Von daher konnte sich die Versuchsanweisung zunächst darauf beschränken, die Vpn sollten werfen, wie sie es gewohnt waren - allerdings ohne Gegner, also im Alleingang - und zwischen den einzelnen Würfen und Wurfserien möglichst viel verbalisieren

(a) zu eigenen Zielsetzungen hinsichtlich ihrer Leistungen;
(b) zur Beurteilung ihrer Leistungen;
(c) zur eigenen Zufriedenheit bzw. Unzufriedenheit bezüglich der jeweils erreichten Leistungen.

Von Fall zu Fall wurden während des Versuchs Variationen eingeführt, z.B. die Vpn gebeten, zur Beurteilung ihrer Leistungen die Begriffe "gut", "mittel" und "schlecht" zu verwenden.

Auf die unmittelbare Wettkampfsituation wurde bewußt verzichtet. Wir wollten ja hier das individuelle Leistungsverhalten der einzelnen Vpn untersuchen - wie es sich auf dem Hintergrund ihrer jeweiligen ganz persönlichen Wertungsmaßstäbe darstellt. Dabei wäre durch direkten Partnervergleich eine hier unerwünschte Komponente ins Spiel gekommen.

Die Versuche dauerten meist mehrere Stunden, wurden teilweise auch über mehrere Tage fortgesetzt; alle Äußerungen der Versuchspersonen wurden auf Tonband aufgenommen, daneben wurden Protokolle geführt.

Während dieser Versuche wurden im allgemeinen von jeder Vp 300 - 500 Würfe "absolviert", vereinzelt kamen auch weitaus längere Serien vor.

Nach jedem Versuch bzw. jeder Versuchsreihe wurden alle Vpn im Einzelgespräch interviewt, die einzelnen Versuchsphasen und Reaktionen wurden durchgesprochen, um Verdeutlichungen über das Versuchsverhalten gebeten. Dabei wirkte sich aus-

gesprochen positiv aus, daß das Gros der Versuchspersonen aus dem Kollegenkreis stammte: sie waren im Versuch Vpn wie andere auch, ganz von der Situation gefangen und dem Gesetz des Spiels, der Auseinandersetzung mit der eigenen Leistung unterworfen; nach dem Versuch hingegen konnten sie als Fachkollegen angesprochen, als Diskussionspartner herangezogen werden. Mitunter ergaben sich aus dieser Doppelfunktion heraus recht bezeichnende Diskrepanzen, die ganz im Sinne der Versuchsanordnung lagen, auf diese Weise wahrscheinlich besonders zur Verdeutlichung beitrugen - z.B. wenn der "Fachkollege" nach dem Versuch einzelne Aspekte seines Verhaltens als "Versuchsperson" nicht mehr so recht wahr haben wollte.

Als grundsätzliche Schwierigkeit stellte sich in den Versuchen heraus, daß die Versuchspersonen generell nicht sehr verbalisierungsfreudig waren. Dies lag jedoch ganz offensichtlich nicht an mangelndem guten Willen, sondern an der Spielsituation: meist waren die Vpn so von der Faszination mitgerissen, daß alle jeweils verfügbaren Pfeile bei voller Konzentration abgeworfen wurden, für Zwischenbemerkungen blieb hier oft einfach keine Zeit und kein Platz. Diesem Dilemma konnte in einigen Fällen dadurch abgeholfen werden, daß spontane Ausrufe wie "Aaah!" oder "Och.", teilweise auch kaum-verbale Äußerungen, die schon eher als Stöhnlaute zu bezeichnen sind, z.B. den Begriffen "gut" oder "schlecht" zugeordnet wurden - in den meisten Fällen war dies durchaus eindeutig. Daneben mußte aber dann zuweilen auch der Versuch unterbrochen und die Vp um eine Beurteilung gebeten werden, was zweifellos jeweils die Situation ein wenig beeinträchtigte: es mußten Bewußtseinsprozesse eingeschaltet werden, die zumindest die Gefahr in sich bargen, daß Unmittelbarkeit und Spontaneität in der Auseinandersetzung mit der eigenen Leistung gemindert wurden.

3.3. Ergebnisse

Die im folgenden berichteten Befunde dieser Versuchsreihe können nicht in tabellarischer oder ähnlicher Form zusammengefaßt wiedergegeben werden, weil die einzelnen Versuche im Sinne der oben dargelegten Planung recht unterschiedlich nuanciert durchgeführt wurden, um möglichst viel - und dies eingehend - vom Spektrum des zu untersuchenden Leistungsverhaltens zu erfassen.
Stattdessen wird hier dargestellt, was sich nach der weitestgehend qualitativen Analyse der Versuchsprotokolle in Verbindung mit den Erfahrungsberichten der Vpn und den jeweils anschließenden Versuchsdiskussionen aus dieser Untersuchungsreihe ergibt. Insofern wird hier auch Zitaten ein relativ breiter Raum eingeräumt, weil sie vor allem in diesem Zusammenhang die anschaulichsten Belege sind.
Im ersten Teil dieses Absatzes wird dabei eher auf diejenigen Aspekte eingegangen, die von der Art der Versuchsanordnung her bestimmt sind bzw. in ganz allgemeinem Sinne für den Bezugssystemcharakter unserer Daten typisch. Die eigentliche Dynamik im Umgang mit den individuellen Bezugssystemen wird dann in Zusammenhang mit dem Erleben von Erfolg oder Mißerfolg anschließend behandelt.

3.3.1. Zur Versuchsanordnung und zu allgemeinen Bezugssystem-Charakteristika

a. Ein erheblicher Teil des Bezugsrahmens für den phänomenalen Wert von Würfen ist von der Struktur der Wurfscheibe her gegeben. Durch die auch auf Wurfentfernung deutlich sichtbaren Ringe und Ringgrenzen erfolgt eine spontane Zuordnung von Begriffen zu Werten, die sich in erfolgsunabhängigen Beurteilungen ausdrückt.
Dies gilt auf der einen Seite für relativ ungeübte Versuchspersonen, die etwa bei ihren ersten Zielsetzungen den 50-Ring der Scheibe als mittleren Wert favorisie-

ren, weil sie ihm einen mittleren Schwierigkeitsgrad zusprechen. Nachfolgende Wertungen und Angaben über die Zufriedenheit mit der eigenen Leistung werden entsprechend auf diese Ausgangssituation bezogen.
Auf der anderen Seite erliegen aber auch geübtere Vpn dieser so ausgeprägten Prägnanz der Wurfscheibe. Sobald in einem Versuch relative Leistungskonstanz erreicht war, lag es für einige Vpn offensichtlich nur allzu nahe, z.B. prinzipiell 50-Würfe als "mittel" zu werten, 10- und 25-Würfe als "schlecht" und 75- und 100-Würfe als "gut". Bei Einstufungen nach fünf Kategorien drängte sich teilweise eine derartig schematische Klassifizierung sogar noch eher auf. Dieses Urteilsverhalten impliziert auch Korrekturen, wie etwa das folgende Beispiel zeigt:

> Der Pfeil steckt zwischen 50 und 75. Die Vp (D) sieht eine 75 und urteilt: "gut - nein, mittel ist das", nachdem sie sich an der Scheibe überzeugt hat, daß es in Wirklichkeit eine 50 war.

Beurteilt werden hier also imgrunde nicht so sehr durch eigene gezielte Handlungen erreichte Leistungen, sondern eher durch relative Zufallsschwankungen gegebene Leistungsprodukte.

b. Auch die farbliche Gestaltung der Wurfscheibe hat das Leistungsverhalten und darüber hinaus auch die Wertungen der Leistungen ganz augenscheinlich beeinflußt: Wahrscheinlich durch seine außerordentliche Prägnanz kann der gelbe 75-Ring zwischen der schwarzen 50 und der ebenfalls schwarzen 100 zum eigentlich anvisierten Ziel werden, das zusätzlich durch seine scharfen Grenzen Leistungsstress und Unsicherheit verursacht.
So kann es geschehen, daß einzelne Versuchspersonen tatsächlich weder den 100-Kreis als eigentlichen Mittelpunkt der Scheibe und als Bestleistung anstreben noch den 50-Ring als Mittel-Ring (in zweierlei Sinn), sondern den dazwischen liegenden besonders prägnanten 75-Ring.

Besonders deutlich wird dies etwa bei der Vp N: Sie setzt sich zunächst die 50 als Ziel, erhöht nach Erfolg auf 75 - und stellt dann einige Zeit später nach Mißerfolg plötzlich fest, sie habe die ganze Zeit zuvor die 75 als Ziel auch anvisiert; in Zukunft wolle sie versuchen, möglichst gut zu werfen und demzufolge die 100 als Scheibenmittelpunkt anstreben. Vp N merkt hier auch durchaus selbst, daß da etwas nicht so ganz stimmt, sie zieht sich - mindestens subjektiv - "aus der Affäre", indem sie zwischen dem "Idealziel" 100 und dem "Realziel" 50 unterscheidet - beides zunächst Ziele im Sinne der Anspruchsniveausetzung, die Wurfziele im Sinne von Fixationspunkten sind damit nicht unbedingt identisch.

c. Daneben gibt es bei Summation mehrerer Wurf-Werte - zu der einige Vpn spontan tendierten, vermutlich weil sich einzelne Fehlwürfe so eher ausgleichen, relativieren lassen - offensichtlich besondere Prägnanz-Strukturen, die ebenfalls letztlich durch die Eigenart der Wurfscheibe vorgegeben sind. Gemeint sind hier die besonderen Marken, die durch unser gebräuchliches Dezimalsystem nahegelegt werden: Werte über 400 (aus 6 Würfen) sind "gut", unter 300 "schlecht", unter 200 "sehr schlecht" - auch dies relativ durchgehend und relativ unabhängig von Erfolg oder Mißerfolg. Ganz offenkundig haben diese 100-Vielfachen eine derart starke Stellung in unseren subjektiven Zahlenräumen, daß sie die Tendenz haben, ähnliche Ordnungen zu assimilieren - und im vorliegenden Fall war die Struktur der Wurfscheibe dementsprechend.

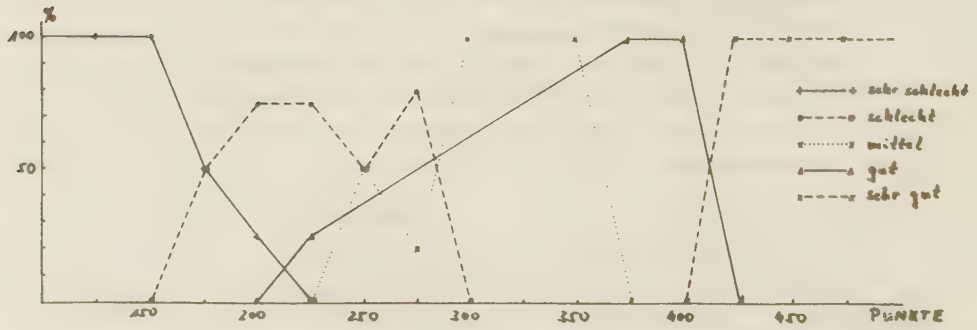
Über diese ganz offenkundigen Charakteristika für das Vorhandensein von Bezugssystemen zur Beurteilung der eigenen Leistungen, die hier von der Art und Struktur der Wurfscheibe, also vom Versuchsmaterial her nahegelegt werden, über diese Aspekte hinaus gibt es aber noch eine ganze Reihe von Momen-

ten, die im engeren Sinne Struktur und Dynamik der Bezugssysteme für die eigene Leistung betreffen, solche nämlich, die weit eher in der Person, im einzelnen Individuum begründet liegen als jene. Aus diesem Bereich lassen sich nach unseren Versuchen vor allem die folgenden Aussagen machen:

d. Gerade im zuletzt angesprochenen Fall der Wert-Summation aus jeweils sechs Würfen (was durch die sechs Pfeile nahegelegt wird) zeigen sich recht deutlich hinsichtlich der Verteilungen der Beurteilungen Tendenzen, wie sie nach der WITTEschen Bezugssystemtheorie und nach den sonst schon vorliegenden Ergebnissen zu erwarten sind.
Ein besonders anschauliches Beispiel hierfür bieten die Daten der Versuchsperson D. Sie urteilt von sich aus spontan nach fünf Kategorien; verwendet werden dabei verschiedene, aber praktisch ausschließlich eindeutig zuzuordnende Begriffe wie z.B.

sehr gut, das ist doch wohl Klasse
gut, ganz gut
mittel, mittelmäßig
schlecht, ziemlich schlecht
sehr schlecht, unheimlich schlecht

Die Wurf-Ergebnisse und ihre Bewertungen von dieser Vp sind in der folgenden Darstellung wiedergegeben:



Darst.1: Bewertung von Pfeilwurf-Ergebnissen im Fünf-Kategorien-Versuch: relative Häufigkeiten

Aus der Verteilung der Ergebnisse von D wird gleichzeitig deutlich, daß der obere Teil des Leistungsspektrums - etwa **alle** Summenwerte ab 300 - keinerlei Unsicherheit in der Beurteilung unterliegt, die Urteilsklassen "mittel", "gut" und "sehr gut" sind hier eindeutig.
Die Vermutung, dies liege daran, daß D nur selten Werte ab 300 erreichte und deshalb gar nicht in der Beurteilung schwanken konnte, läßt sich nicht bestätigen, denn die Werte unter 300 machen nur unwesentlich mehr als die Hälfte von D's Wurfergebnissen aus (54%). Dennoch dürfte kaum von der Hand zu weisen sein, daß hier die Übung eine erhebliche Rolle gespielt hat, die den "unteren" Leistungsbereich (bezogen auf diese Aufgabe) für D interessanter machte, was sich in ihrem für diesen Bereich besonders wechselnden Urteilsverhalten ausdrückt. In diesem Falle ergänzen sich die Resultate aus der Datenanalyse auch mit unseren Beobachtungen im Versuch besonders gut.

e. Auch in diesem Untersuchungsfeld scheint es - im übertragenen Sinne - das Phänomen der "Bleistifthaftigkeit" (WITTE mündlich) zu geben: ein Bleistift muß z.B. schon eine bestimmte Mindestlänge haben, um wenigstens als "sehr, sehr kurzer" Bleistift bezeichnet werden zu können (entsprechendes gilt für den Maximalbereich, irgendwann wäre der Bleistift kein Bleistift mehr).
Analog gibt es Leistungen - auch in unseren Versuchen deutlich feststellbar - die so "schlecht" sind, daß sie schon nicht mehr als schlecht bezeichnet werden; sie fallen vollständig aus dem Rahmen, werden nicht als Mißerfolg erlebt und tangieren die Eigeneinschätzung nicht im geringsten (wenn sie sich nicht häufen!).
Einen sehr deutlichen Hinweis hierauf gibt uns etwa die Vp W, einer der geübtesten Mitspieler: "Unter 300 ist schlecht, das stört mich - aber 160, das macht mir gar nichts; ich

weiß, so schlecht bin ich nicht." Die Wurfleistung 160 (mit 6 Würfen) ist also für W nicht im mindesten Grund zur Unzufriedenheit, kein Grund zur Sorge, wie es wahrscheinlich der "bessere" Wert 260 wäre - 160 liegt vollkommen außerhalb des Bezugsrahmens.
Eine entsprechende Obergrenze bei der Einschätzung der eigenen Leistungen konnten wir in diesen Versuchen nicht feststellen, obwohl durchaus - und dies gar nicht so selten - Ergebnisse vorkamen, die im Vergleich zu den anderen so "gut" waren, daß man entsprechende Vermutungen hätte anstellen können. Denkbar wäre auch, daß derartige "Erfolge" mindestens subjektiv deutlich seltener sind als entsprechende "Mißerfolge" und so andere Stellenwerte haben.

f. Darüber hinaus scheint der Aufbau von Bezugssystemen bei dieser Art Aufgabe im wesentlichen auf zwei Wegen vor sich zu gehen, soweit man den Erfahrungsberichten der Vpn und den Versuchsbeobachtungen entnehmen kann:
Zunächst einmal gibt es eine Reihe von Versuchspersonen, die - wie bereits oben in anderem Zusammenhang erwähnt - den 50-Ring der Scheibe als mittleren Ring favorisieren, weil sie ihm einen mittleren Schwierigkeitsgrad zusprechen. Sie erleben alle Würfe, die mindestens diesen Ring treffen, als Erfolge und bezeichnen sie als "zufrieden"-stellend oder "gut". Umgekehrt werden alle schlechteren Würfe als Mißerfolge erlebt und als "schlecht" bezeichnet.
Andere Versuchspersonen gehen sozusagen von außen an die Scheibe heran: so will z.B. Vp N "wenigstens die Scheibe treffen", wäre von daher also mit jedem Treffer zufrieden und würde nur die sprichwörtlichen "Fahrkarten" als Mißerfolge erleben. Allerdings ist gerade diese Vp gleichzeitig ein Beispiel dafür, daß es bei dieser Art der anfänglichen Zielsetzungen durchaus erhebliche Nuancen geben kann: sie will nicht nur immer wenigstens die Scheibe treffen, son-

dern gleichzeitig "dabei mit jedem Wurf mindestens 25 Punkte erreichen" - nicht etwa nur 10, wie es dem äußersten Ring entspräche! Ganz niedrig darf also auch bei dieser Gruppe von Versuchspersonen selbst das Anfangsziel nicht gesetzt werden; wenigstens eine Stufe muß phänomenal erklommen sein, um den Erfolg als solchen erleben zu können - allerdings ist die Mißerfolgswahrscheinlichkeit recht gering, zumindest im Vergleich zu der zuerst genannten Gruppe von Vpn, die nicht von einem Mindestwert, sondern von einem subjektiv mittleren Schwierigkeitsgrad ausgehen. Den Fall einer minimalen Zielsetzung (10 für jeden Wurf oder 60 als Summenwert) haben wir im übrigen in unseren gesamten Versuchen nicht ein einziges Mal erlebt.

g. Weiterhin können wir nach unseren Befunden feststellen, daß die meisten Vpn mit mehreren Bezugssystemen gleichzeitig "arbeiten", die etwa alle in der Art der Aufgabe, in der Aufgabenstruktur ihre Basis haben können. Unterschiedlich ist an ihnen einzig, daß sie sich in verschieden langen Wurfserien manifestieren und von daher auch verschiedene Bezeichnungen tragen können. So gibt es z.B. ein System für die "Tagesform" (30 Würfe bei Vp W), ein weiteres für den "Grad der Geübtheit" (Summenwert aus 150 Würfen) - dies alles jeweils neben den Maßstäben für Einzelergebnisse und Summenwerte z.B. aus sechs Würfen (weil sechs Pfeile zur Verfügung standen).
An dieser Stelle mag die Erwähnung der gleichzeitigen Existenz mehrerer Bezugssysteme genügen; auf den individuellen Umgang mit ihnen wie auch auf die Beurteilungsmaßstäbe sozialer Genese wird im Zusammenhang mit der Thematisierung des Erlebens von Erfolg oder Mißerfolg näher einzugehen sein.

3.3.2. Zur Auseinandersetzung mit Erfolg und Mißerfolg

Eigentlich zum Tragen kommen die Strukturprobleme der individuellen Bezugssysteme, die Aspekte ihres Aufbaus und ihrer Veränderungen jedoch erst vor dem Hintergrund von Erfolg und Mißerfolg, diesen für unser Leistungsverhalten entscheidenden Erlebnissen. Erst in der hier gespiegelten Auseinandersetzung der Vpn mit der eigenen Leistung wird die Dynamik der Bezugssysteme auf diesem Gebiet deutlich.

a. Verschiebungen der Zielsetzungen und Beurteilungsmaßstäbe innerhalb eines einzigen Bezugssystems stellen sich als noch leidlich übersichtlich dar: in genau definierten Einheiten - z.B. 50 Punkte für den Summenwert aus sechs Würfen - wird je nach Erfolg oder Mißerfolg das Anspruchsniveau gehoben oder gesenkt. Dabei liegt aufgrund des vorliegenden Materials die Vermutung nahe, daß die Änderungsschwelle für den Fall der Erfolgsverschiebung sich von derjenigen für die Mißerfolgsverschiebung erheblich unterscheidet: Genügt für die Erhöhung des Anspruchsniveaus im allgemeinen ein einziger Erfolg (Übertreffen des zum Ziel gesetzten Punktwertes), so sind für die Verminderung mindestens drei bis vier Mißerfolge notwendig, davon mindestens ein ganz massiver (z.B. beim Summenwert aus sechs Würfen ein Unterschreiten des Zielwertes um mindestens 100 Punkte). Wie groß dann dabei die Diskrepanz zwischen Soll-Wert und Ist-Wert auch ist - Erhöhung oder Verminderung des Zielwertes geschehen in den definierten Einheiten (z.B. im Fall des angesprochenen Summenwertes von je 50 Punkten). Dieses Verhalten läßt darauf schließen, daß Erfolge zwar schnell und bereitwillig als "Erfolge" erlebt werden, Mißerfolge jedoch zunächst einmal nicht als "Mißerfolg", sondern als "Ausbleiben des Erfolges", als Verzögerung. Erst dann, wenn sich der Mißerfolg, d.h. das Nicht-Er-

reichen des Zieles, nur noch äußerst schwer übersehen läßt, wird der Anspruch gemindert, aber auch jetzt nur in relativ kleinen Schritten. - Wir sehen dies wohl nicht zu Unrecht als Parallele zu unserem oben schon berichteten Ergebnis an, daß wir zwar das Vorhandensein einer unteren Systemgrenze bezüglich der Beurteilung der eigenen Leistung feststellen konnten, nicht aber die Existenz ihres oberen Äquivalents. Auch an dieser Stelle soll aber daneben auf die Erklärungsmöglichkeit hingewiesen werden, daß - wenigstens subjektiv - die Erfolgswahrscheinlichkeit in dieser Versuchsanordnung geringer ist als die Mißerfolgswahrscheinlichkeit, was zu verschiedenen Verhaltensweisen führen kann.

b. Auch darüber hinaus sind für die Frage nach "Struktur, Dynamik und Genese von Bezugssystemen" (WITTE 1960a) bei der Beurteilung eigener Leistungen die Mißerfolgsreaktionen unserer Versuchspersonen um einiges vielschichtiger und aufschlußreicher als die Verhaltensweisen nach Erfolg. Wird nach Erfolg im allgemeinen das Anspruchsniveau nur innerhalb eines Bezugssystems heraufgesetzt, so ist es im umgekehrten Falle mit der Herabsetzung nur über eine begrenzte Zeit getan. Für alle Vpn scheint es einen kritischen Punkt zu geben, der nach unten hin nicht überschritten werden darf - das Anspruchsniveau kann nicht beliebig niedrig angesetzt werden.
Dergleichen hatte sich oben bereits angedeutet: Vp N wollte mit jedem Wurf wenigstens die Scheibe treffen und dabei jeweils mindestens 25 Punkte erreichen, nicht etwa nur 10. Gemeint ist hier etwas anderes als die oben zitierte "Bleistifthaftigkeit", die für die Systemgrenze steht. Hatte es sich dort um Fälle gehandelt, wo die Leistung klar außerhalb des Bezugssystems liegt, wo ihr Ergebnis nicht etwa als Mißerfolg empfunden wird - so geht es hier um Leistungs-

resultate, die eben nicht außerhalb des Gütemaßstabes liegen, sondern innerhalb des Systems so schwer wiegen, daß sie die Vp zu ernsthaften Maßnahmen zwingen.
Wir konnten beobachten, daß bei Erreichen dieses kritischen Punktes eine Art Flucht in andere Bezugssysteme stattfindet. Unsere Versuchspersonen definierten relativ plötzlich - offen oder versteckt, deutlich oder beschönigend - die Aufgabenstellung um, wodurch nunmehr andere Gütemaßstäbe relevant wurden.
Dies kann offenkundig auf derart mannigfache Art geschehen, daß das Spektrum wohl kaum je ganz auszuloten sein wird. Wir schildern darum hier vergleichsweise ausführlich einige Verhaltensbeispiele, um damit gleichzeitig auch wenigstens ein wenig von der Variabilität und Findigkeit in der Auseinandersetzung mit der eigenen Leistung zu dokumentieren.

(1) Einer der naheliegendsten und am häufigsten eingeschlagenen Wege beim Bezugssystemwechsel ist der über eine Erhöhung der Wurfanzahl. Dies gilt sowohl für weniger erfahrene als auch für ausgesprochen routinierte Vpn und kann sich in verschiedener Weise ausdrücken:
Z.B. gibt Vp B an, sie wolle mit jedem Wurf mindestens 50 Punkte erreichen, <u>also</u> mit sechs Würfen 300. Ein möglicher Mißerfolg wird damit bereits antizipatorisch relativiert, indem Kompensationsmöglichkeiten eingebaut werden. Sollte wirklich ein Wurf schlechter sein als 50, so kann er durch eine 75 oder gar 100 wieder wettgemacht werden - imgrunde wird hier bereits mit zwei Gütemaßstäben, mit zwei Bezugssystemen gleichzeitig gearbeitet; ein Mißerfolg in dem einen läßt einen möglichen Erfolg in dem anderen relativ unberührt.
Diese Vorgehensweise läßt sich fast beliebig erweitern von 6 Würfen auf 30, auf 150 usw., es lassen sich verschiedene Bezeichnungen kreieren wie die oben bereits zitierten "Grad der Geübtheit" oder "Tagesform" etwa. Zwischen den verschie-

denen Bezugswerten manövrieren die Vpn relativ leicht hin und her; es müßte schon ziemlich viel passieren, wollten sie nicht wenigstens in einem der Systeme Erfolg haben. Vollkommen problemlos - aus der Sicht der Versuchsperson - ist diese Verfahrensweise dennoch nicht, birgt sie in sich doch gleichzeitig mit den Kompensationsmöglichkeiten die Gefahr des Vabanque-Spiels - wenn nämlich zu lange auf Ausgleich gehofft und gesetzt wird. In diesem Falle ist der Mißerfolg am Ende möglicherweise ein totaler, der praktisch durch nichts mehr abgefangen werden kann - wie beim Roulette-Spieler, der vielleicht auch beim allerletzten Jeton noch vergeblich hofft, jetzt müsse der Umschlag des Spiellaufes zum Gewinn hin aber wirklich kommen. Wir können hier nur vermuten - da wir diese Extremgrenzen nicht austesten konnten - , daß bei diesem totalen Mißerfolg ebenfalls die Systemgrenzen überschritten werden: der "Mißerfolg" wäre dann keiner mehr, sondern so etwas wie ein Schicksalsschlag (wobei wir hinter den Roulette-Vergleich wenigstens ein Fragezeichen setzen, was Erfolg oder Mißerfolg aufgrund eigener Leistung angeht).

(2) Ein weiterer Weg zur Beeinflussung des Erfolgs- bzw. Mißerfolgserlebens durch Bezugssystem-Wechsel liegt in einer semantischen Differenzierung.
So hatte sich z.B. eine unserer Versuchspersonen das Ziel gesetzt, mit jedem Wurf mindestens 50 Punkte zu erreichen. Nach einer längeren Wurfserie mit einer Erfolgsquote von etwa 50% erhöht sie das Ziel auf 75, weil sie glaubt, zu lässig zu werfen und sich so eher anzustrengen. Nach einer ausgesprochenen Mißerfolgsserie - nur 20% der Würfe waren im Sinne der Zielsetzung erfolgreich - wird das Leistungsziel wieder vermindert "nicht weil ich meine, ich könnte die 75 nicht treffen, sondern weil ich glaube, daß ich etwas mehr Spielraum brauche ... Ich habe mich mit den 75 so eingeengt, und je enger der Spielraum ist, desto intensi-

ver bemühe ich mich, nun diesen kleinen Punkt zu treffen." Nun, niemand hatte verlangt, daß der 75-Ring zum Wurf-Ziel erklärt werden sollte. Es hat den Anschein, als bemühe sich unsere Vp hier nicht eben um sprachliche Klarheit, zumal sie nach einer weiteren Reihe von Würfen fortfährt: "100 sind Idealziel, 50 sind Realziel; 100 peile ich an, mit 50 bin ich zufrieden." Diese und ähnliche Äußerungen legen den Schluß nahe, daß "anpeilen" hier durchaus doppeldeutig gemeint ist, nämlich einerseits als anvisieren im Sinne des Leistungsvollzuges, andererseits als anstreben im Sinne der Leistungsbewertung - ein Beurteilungsaspekt, der sich übrigens auch im anschließenden Interview nicht aufklären ließ. Ähnliches kam bei anderen Versuchspersonen vor, gerade bei dieser Vp jedoch verstärkte sich der Eindruck einer "bewußten" (?) Undeutlichkeit noch im weiteren Verlauf des Versuches, als sie schließlich den 50-Ring als Mindestwert angibt, den sie erreichen will. Nach dieser Zielsetzung wären also alle Würfe unter 50 Punkten Mißerfolge. Um dies jedoch nicht zu deutlich werden zu lassen - es passiert in diesem Versuchsstadium immer häufiger - wird dann aber der 50-Ring ausdrücklich zum eigentlichen Ziel erklärt: jetzt ist 25 so gut (oder schlecht) wie 75, 10 so gut wie 100 - es kann nicht mehr kleinere oder größere Abweichungen vom Mittelpunkt der Scheibe geben, sondern nur noch kleinere Differenzen nach beiden Seiten hin. Außerdem ist eine Abweichung zur 100, zum Mittelpunkt der Scheibe hin wohl auch nicht gerade ein sehr "schlimmer" Mißerfolg!

Diese Versuchsperson - und sie mag hier nur als besonders plastisches Beispiel dieses Phänomens dienen - manipuliert also durch derartige Umdefinierungen während des Spiels imgrunde ziemlich ungeniert die Wahrscheinlichkeiten (soweit man sie theoretisch erfassen kann) von Erfolg und Mißerfolg zu ihren Gunsten - denn natürlich war die Versuchsanweisung

gewesen, so gut wie möglich zu werfen, und daran war auch während des Versuches nichts geändert worden.

(3) Einen noch erheblich aparteren Weg zur Manipulation der theoretischen Wahrscheinlichkeiten von Erfolg und Mißerfolg fand die Versuchsperson D über die Farbe der Pfeile. Sie hatte nach einer längeren Mißerfolgsserie plötzlich die Idee, die Wurfpfeile von vornherein nach ihrer Farbe zu bewerten (z.B. rot=gut, gelb=schlecht) und so ihren phänomenalen Mißerfolgsbereich, ihre Mißerfolgswahrscheinlichkeit zu verkleinern: Ein guter Wurf mit gutem Pfeil ist natürlich gut - ein schlechter Wurf mit schlechtem Pfeil aber mitnichten schlecht, sondern "selbstverständlich", er ist nach ganz anderen Maßstäben zu beurteilen.
Ein guter Wurf mit schlechtem Pfeil ist eine angenehme Überraschung, ein ungeplanter Erfolg - Mißerfolg ist praktisch nur noch ein schlechter Wurf mit gutem Pfeil. Diese Versuchsperson schafft es so, die theoretische Erfolgswahrscheinlichkeit von 50% : 50% zu erhöhen auf 75% : 25%.
Die Leistungsziele brauchen auf diese Art noch nicht einmal verändert zu werden; die Maßstäbe zur Bewertung der eigenen Leistung können aufrecht erhalten werden, indem man im Versagensfalle auf andere Ebenen ausweicht.

(4) Letztes - und das nicht nur in dieser Kasuistik - Ausweich-Bezugssystem nach längeren Mißerfolgsserien, wenn die Leistungsziele auf den verschiedenen Ebenen wirklich nicht mehr gesenkt werden können, ohne zu sehr persönlich zu tangieren, ist schließlich das des Partnervergleichs. Alle Versuchspersonen forderten irgendwann während der Versuche den Versuchsleiter zum Mitspielen auf (was aus den oben genannten Gründen stets abgelehnt wurde). Die am Ende der Versuche stets durchgeführten Explorationen ergaben zu diesem

Aspekt, daß auch hier ganz offenkundig ein Ausweg in ganz andere Bezugssysteme, in ganz andere Gütemaßstäbe angestrebt wird. Dies kann im wesentlichen zwei Ausprägungsformen aufweisen:
Auf der einen Seite können die Vpn nach einer längeren Mißerfolgsserie - und von diesen Situationen ist ja hier in erster Linie die Rede - nur noch auf Erfolg durch einen Sieg über den Spielpartner hoffen, "viel schlimmer kann es nicht mehr werden, höchstens besser". Sollte aller Hoffnung zum Trotz der Partner aber gewinnen, so wäre dies nicht unbedingt ein Mißerfolg (gegenüber selbstgesteckten Zielen bezüglich der individuellen Leistung), sondern der andere war eben besser, geübter o.dgl. - und das ist schließlich keine Schande, dagegen kann man nichts machen.
Daneben kann der Wunsch unserer Vpn nach Erfolg durch Sieg über den Spielpartner aber auch noch in einer anderen Nuance wichtig sein: sie wollen weiterspielen trotz beträchtlicher Mißerfolge, ihre sämtlichen individuellen Gütemaßstäbe können endgültig und total über Bord gehen - es genügt ihnen, wenn sie am Ende besser sind als, z.B. als der Versuchsleiter. Sie flüchten also gleichsam in die Relativität, die bisherigen Absolut-Beurteilungen der eigenen Leistung werden aufgegeben.

3.4. Zusammenfassung

Wir haben gesehen, daß bei der Auseinandersetzung mit der eigenen Leistung - dargestellt am Beispiel des Dart-Spiels - eine erhebliche Dynamik hinsichtlich der Bewertungsmaßstäbe vorliegt. Bei ihrem "Bestreben, die eigene Tüchtigkeit ... zu steigern oder möglichst hoch zu halten" (HECKHAUSEN 1965), erleben unsere Versuchspersonen Erfolg und Mißerfolg ganz augenscheinlich ausgesprochen intensiv. Es sind deutlich Anspruchsniveau-Setzungen und -Verschiebungen feststellbar, die sich auf ganz verschiedene Gütemaßstäbe beziehen können, näm-

lich auf solche, die (a) von der Aufgabenstellung, von der Art etwa des Spiels her; (b) von der eigenen Erfahrung her; oder schließlich (c) sozial, d.h. durch andere, als relevant angesehene Personen determiniert sein können (vgl. u.a. HECKHAUSEN 1969).

Auf wie vielen Ebenen sich dabei die Beurteilung der eigenen Leistung abspielen kann, kann durch die angeführten Beispiele lediglich angedeutet werden. Vor allem im Falle des Mißerfolges scheint der individuelle Erfindungsreichtum - sofern man hier vorwiegend kognitiv argumentieren will - einigermaßen unerschöpflich zu sein: kommt es auf der einen Seite schon wesentlich seltener zu einer Verschiebung des Nullpunktes der angelegten Gütemaßstäbe nach "unten" als nach "oben", so liegt daneben der Sprung in ein ganz anderes System oft ebenso nahe, spätestens als letzte Rettung vor der "Demaskierung" zu hoch gesteckter Ziele (zu hoch zumindest für die augenblickliche Leistungsfähigkeit) wird er vorgenommen. Ein erheblicher Teil der Bezugssystem-Dynamik auf dem Gebiete der Leistungsmotivation scheint sich also zwischen verschiedenen Systemen abzuspielen, da wir als typisches Mißerfolgsverhalten neben dem Senken des Anspruchsniveaus eben die geschilderten Wechsel in andere Systeme feststellen. Wir stoßen damit hier auf ein Phänomen, das in der bisherigen Bezugssystemforschung, bei der es primär um physikalisch nachprüfbare Sachverhalte (bzw. um Mannigfaltigkeiten, die sich über physikalische Metriken anordnen liessen) ging, ohne Belang war und auch bleiben konnte: Welche Versuchsperson ist z.B. schon in einem Versuch zur Beurteilung von Gewichten plötzlich vom Urteil "leicht" übergegangen zu "fühlt sich angenehm an" oder ähnlichem? Hier bei der Bewertung der eigenen Leistung finden wir also insoweit eine erheblich andere Situation vor.

In dieser Arbeit soll es aber nicht so sehr um die Dynamik

des Mißerfolgsverhaltens und damit um den Wechsel zwischen verschiedenen Bezugssystemen gehen (wenngleich wir dergleichen, wie sich herausgestellt hat, gar nicht außer Acht lassen können) - wir wollen hier in erster Linie versuchen, der Dynamik und den Strukturen innerhalb der einzelnen Gütemaßstäbe nachzugehen, wollen sehen, ob im Leistungsbereich ebenfalls die von WITTE auf der Basis intersensorieller Versuche formulierten Gesetzmäßigkeiten gelten.
Unsere bisherigen Ergebnisse haben dazu bereits gezeigt, wenigstens in einer deutlichen Tendenz, daß auch für die individuellen Einschätzungen der eigenen Leistungen WITTEs Theoreme gelten. Wir wollen daher im folgenden - ganz im Sinne der oben dargelegten Versuchsplanung - versuchen, auf der Basis der skizzierten Bezugssystem-Kasuistik nähere Aussagen zur Struktur und Dynamik innerhalb einzelner Wertungsmaßstäbe zu machen, die für das individuelle Leistungsverhalten relevant sind.

4. Struktur und Dynamik individueller Leistungsbeurteilungen am Beispiel von Reaktionszeitversuchen

4.1. Versuchsplanung

Nach aller Erfahrung gehören Reaktionszeit-Meßgeräte zu den Apparaturen, denen der Vorübergehende oft nur schwer widerstehen kann. Die meisten von uns reizt es immer wieder, z. B. zu sehen, "wie schnell man ist" und ob es nicht doch noch schneller geht.
Nicht zuletzt aus diesem Grunde wählten wir für unsere weitergehende Analyse der Bezugssystem-Strukturen auf dem Gebiete des Leistungsverhaltens die Messung der einfachen Reaktionszeit (akustisch) aus - wollten wir doch mit Versuchsanordnungen arbeiten, die den Beweggrund zur "Auseinandersetzung mit Gütemaßstäben" möglichst bereits in sich bargen, sodaß der Motivierung der Vpn - wenn wir schon aus Gründen der Übersichtlichkeit im Laborversuch arbeiten mußten - kein eigens gesetzter Impuls vorausgehen mußte.
Daneben war für die Wahl dieser Versuchsanordnung bestimmend, daß zunächst einmal jeder mit seiner eigenen Reaktionszeit unter den verschiedensten Bedingungen des täglichen Lebens mannigfaltige Erfahrungen hat. Wir können also davon ausgehen, daß es sich hier um ein mnestisch stabilisiertes Bezugssystem im WITTEschen Verständnis handelt, daß höchstens hinsichtlich speziell dieser Art der Reaktionszeitmessung eine "Einkennung" sensu HELLPACH (1951) stattfinden muß.

Weiterhin liefern Reaktionszeitversuche recht übersichtliche Daten; die Versuchsbedingungen sind bei allem Leistungsanreiz gut konstant zu halten, sodaß den Rahmenbedingungen nicht allzu viel Aufmerksamkeit gewidmet werden muß.
Die "akustische" Reaktionszeit" (im Gegensatz zur "optischen") wurde gewählt, um dem Versuchsleiter in jeder einzelnen Phase

auch während des Protokollierens Rückmeldungen über den Versuchsablauf zu geben. Dies war nun dadurch möglich, daß der Signalton jeweils erst bei Knopfdruck, auf die Reaktion hin verstummte.
Geplant wurden verschiedene Versuchsreihen unter unterschiedlichen Bedingungen, auf die im folgenden Absatz im einzelnen noch näher eingegangen wird. Ziel sollte in all diesen Versuchen sein, daß die Versuchsperson jeweils möglichst schnell durch Knopfdruck auf ein akustisches Signal reagierte und ihre Reaktionszeit anschließend mit den Begriffen gut bzw. schlecht beurteilte. Kriterium sollte dabei in jedem Fall die Zufriedenheit mit der eigenen Leistung im Sinne des Anspruchsnivesus sein.

4.2. Versuchsdurchführung

Es wurden mehrere Versuchsserien durchgeführt. Unterschiedlich war ihnen zunächst einmal die Anzahl der Kategorien, nach denen geurteilt werden sollte. Dabei wurde zwischen zwei und drei Kategorien variiert. Zu den dazu vorgegebenen Begriffen gut/schlecht bzw. gut/mittel/schlecht bürgerten sich sehr schnell die Synonyme zufrieden/unzufrieden (oder nicht zufrieden) ein, worin sich gleichzeitig ausdrückt, daß die Versuchsabsicht - und Versuchsvoraussetzung - von der starken Ich-Beteiligung unserer Vpn verwirklicht war.

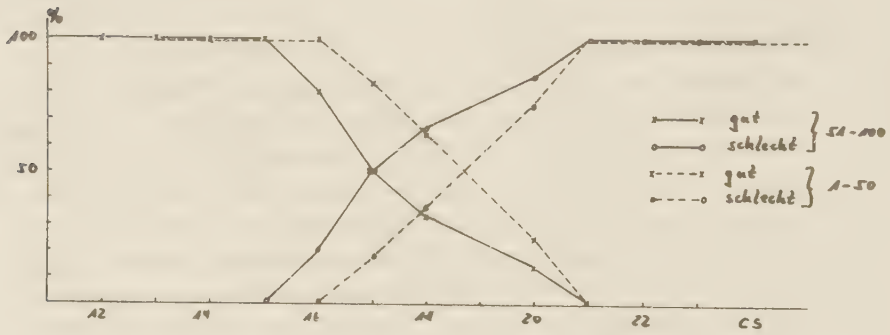
Weiterhin wurden in einigen Versuchsserien unmittelbar die eigenen Reaktionen von den Vpn beurteilt, in anderen die vom Vl nach der Reaktion jeweils mitgeteilten Zeit-Werte. Dieser letztere Weg wurde eingeschlagen zum Versuch der Stabilisierung des zugrunde liegenden Bezugssystems, nachdem sich herausgestellt hatte, daß einige Vpn zur Absolutbeurteilung lediglich aufgrund des Empfindens der eigenen Reaktionszeit nur eingeschränkt imstande waren - sie urteilten stark relativ aufgrund der Erinnerung an den letzten Reiz,

stellten also weitgehend nur Paarvergleiche an. Die Datenanalyse zeigte zwar durch die Häufigkeitsverteilung der Werte bzw. Beurteilungen, daß die Vpn absoluter urteilten als sie selbst meinten, wir gaben ihnen dennoch teilweise die Zeitwerte als Hilfe zur Beurteilung.
Gearbeitet wurde mit einem Reaktionszeit-Meßgerät des Typs Bettendorff T 48, die Meßgenauigkeit betrug Centi-Sekunden (cs). Versuchspersonen waren Psychologie-Studenten älterer Semester, die wiederum aus dem persönlichen Bekanntenkreis gewonnen wurden.

4.3. Ergebnisse

Es zeigte sich insgesamt, daß wir bei der Messung der Reaktionszeit zu Recht von der Annahme ausgegangen waren, es hier mit einem mnestisch stabilisierten System zu tun zu haben. Die Urteile der Vpn waren allgemein sehr sicher und streuten wenig.
Gleichzeitig zeigten die Ergebnisse jedoch auch die grundlegende Problematik dieser Versuchsanordnung, deren Gewichtigkeit sich erst während der Versuche herauskristallisierte und sich auch durch Variationen nicht wesentlich einschränken ließ. Sie besteht hauptsächlich darin, daß das individuelle Leistungsspektrum zu schmal, die zu beurteilende Mannigfaltigkeit der eigenen Reaktionszeiten zu gering ist, als daß sich umfangreichere Bezugssysteme darstellen und ausdifferenzieren können. Insofern ist auch für dynamische Züge nicht viel Raum.
Bei unseren Versuchspersonen kamen Reaktionszeiten vor, die nur über 7 - 25 Intervalle (von Centi-Sekunden) streuten. Setzt man operationalisierend die Systemgrenzen da an, wo nach einer Mindesthäufigkeit der Reaktionszeiten pro Intervall von 2% aller jeweils in einem Versuch vorkommenden Zeiten nur noch Häufigkeiten von unter 2% zu beobachten sind,

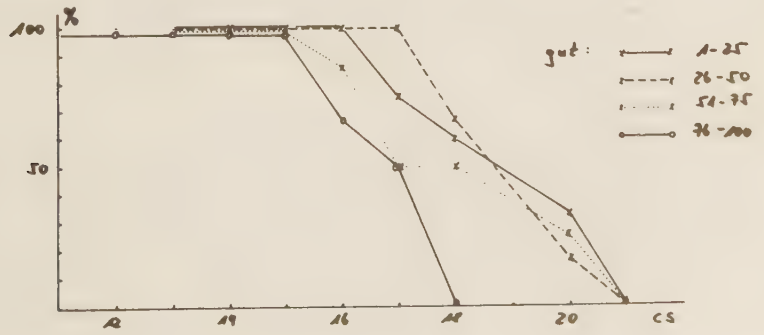
dann wurden gar nur 7 - 14 Intervalle "in Anspruch genommen"; teilweise lagen sogar etwa bei einem Versuch mit nur 9 vorkommenden Zeitklassen 25% der Reaktionszeiten in einer einzigen!
Das Fazit aus dieser Problematik ist, daß sich zuverlässig lediglich Zwei-Kategorien-Systeme feststellen ließen, diese allerdings in aller Deutlichkeit, wie das folgende Beispiel zeigt (Darst.2).



Darst.2: Bewertung von Reaktionszeiten im Zwei-Kategorien-Versuch: relative Häufigkeiten bei 2x50 Durchgg.

Die Darstellung beruht auf den Beurteilungen der Reaktionszeiten aus zweimal 50 Durchgängen der Vp HO, zwischen denen 10 Minuten Pause lagen. Zu beurteilen war die Zufriedenheit mit der Geschwindigkeit der Reaktion aufgrund des eigenen Zeitempfindens der Vp. Die vorkommenden Reaktionszeiten verteilen sich insgesamt zwischen 12cs und 24cs, also über 13 Intervallklassen. Für die ersten 50 Durchgänge beträgt der Systemumfang 11 (14-24cs), für die zweiten beträgt er 12 (12-23cs).
Bereits bei diesen relativ wenigen Urteils-Werten zeigt sich

klar die typische Ogiven-Form der Häufigkeitsverteilungen, die nach der oben angenommenen Bezugssystem-Theorie zu erwarten war. In beiden dargestellten Fällen (Durchgänge 1-50 und 51-100) lassen die "Äste" der Verteilungen auf die GAUSSsche Glockenkurve schließen.
Hinzu kommt die Lage der Zäsuren. Nach den WITTEschen Annahmen (vgl. u.a. WITTE 1971) wären sie zu erwarten gewesen für die ersten 50 Durchgänge bei 19,0cs, für die zweiten bei dem Wert 17,5cs. Tatsächlich liegen die Zäsuren bei 18,7cs bzw. 17,0cs - bereits bei diesen wenigen Daten also eine verblüffend hohe Übereinstimmung.
Deutlich wird aus den in der Darstellung wiedergegebenen Befunden weiterhin neben den genannten Strukturmerkmalen auch die Dynamik des Systems: Bereits die Dichotomisierung der Daten in zweimal 50 Durchgänge genügt, um eine vergleichsweise erhebliche System-Verschiebung sichtbar zu machen: immerhin wandert der Nullpunkt des Systems (hier identisch mit der Zäsur zwischen beiden Beurteilungskategorien) um 1,7cs. Diese Tendenz wird sogar noch eher deutlich, nämlich bereits bei Fraktionierung der gleichen Daten in viermal 25 Durchgänge, wie dies Darstellung 3 zeigt:



Darst.3: Bewertung von Reaktionszeiten im Zwei-Kategorien-Versuch: relative Häufigkeiten bei 4x25 Durchgg.

Eingezeichnet sind hier der besseren Übersicht halber nur die Häufigkeits-Verteilungen der Zufriedenheits-Urteile, diejenigen der Unzufriedenheits-Bewertungen verlaufen analog. Wie bereits angedeutet lassen bereits die Bewertungen der eigenen Reaktionsleistungen aus jeweils nur 25 Durchgängen eine deutliche Dynamik innerhalb des zugrunde liegenden Bezugssystems sichtbar werden: die in der Darstellung nicht eingezeichneten Zäsuren verschieben sich empirisch um insgesamt 1,8cs. Der Ergänzung halber sind sie hier - auch im Vergleich zu den erwarteten Werten - tabellarisch angeführt:

	Durchgang 1-25	26-50	51-75	76-100
erwartet (=e)	18,5	19,0	18,5	16,0
beobachtet (=b)	18,7	18,6	17,5	16,9
Differenz (e-b)	-0,2	+0,4	+1,0	-0,9

(Werte in cs)

Auffällig - und befriedigend im Sinne unserer Hypothesen - sind auch hier wiederum die recht guten Übereinstimmungen zwischen erwarteten und beobachteten Werten, von denen oben bereits die Rede war: Vor allem für die ersten und für die zweiten 25 Durchgänge sind die Diskrepanzen ausgesprochen gering. Sie sind weiterhin - die Tabelle zeigt es - nicht gleichsinnig, was zusätzlich für zufällige Abweichungen spricht.

Wir können so davon ausgehen, daß auf der einen Seite für die Bewertung der eigenen Reaktionsleistungen ein derart stabiles Bezugssystem vorhanden ist, daß es Absolut-Beurteilungen erlaubt, die zudem noch normalverteilt zu sein scheinen. Andererseits ist dieses System flexibel genug, um bei Leistungsänderungen Nullpunkt-Verschiebungen zuzulassen - ohne daß es im übrigen seine Charakteristika ändert.

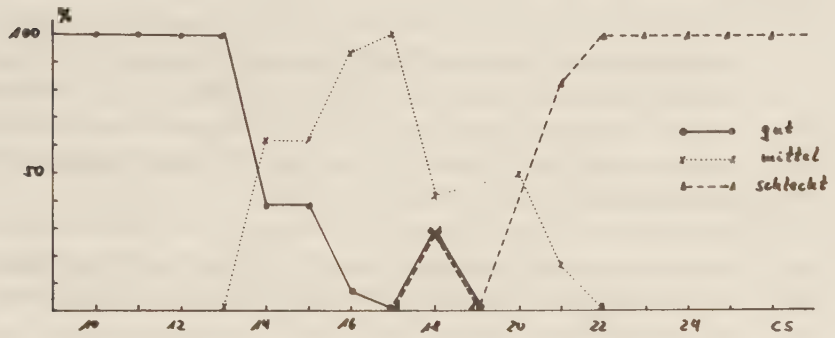
Derartig positive Ergebnisse lassen sich nach unseren Versuchen lediglich für Zwei-Kategorien-Systeme berichten. Vermutlich in erster Linie aufgrund der geringen Mannigfaltigkeit der jeweils zu beurteilenden Reaktionen, aufgrund der intraindividuell geringen Streubreite also, die plausiblerweise nicht zu weiterer kategorialer Differenzierung motiviert und sie auch (in dieser Schwellennähe) kaum möglich macht, zeigen bereits Drei-Kategorien-Versuche kein auch nur annähernd befriedigendes Bild. Dies mag zudem an der Eigenart der akustischen Wahrnehmung liegen, die durch ihren - im Vergleich zur optischen Wahrnehmung - relativ stärkeren Sukzessiv- und schwächeren Simultan-Charakter sich möglicherweise auch bei der Ausdifferenzierung schwerer tut. Längere Versuchsserien, die diesem Problem eventuell hätten abhelfen können - vermutlich wäre allerdings der Leistungsspielraum noch kleiner geworden -, dürften aufgrund der relativ einfach-strukturierten und gleichförmigen Aufgabe kaum möglich sein. Wir führten - auch im Selbstversuch - bis zu 200 Durchgängen durch, was - bei allem Leistungsehrgeiz unserer Versuchspersonen - ziemlich an der Grenze des Erträglichen lag, an der geringen Streuung jedoch kaum etwas änderte, von "Ausrutschern" abgesehen, die in jedem Falle ausserhalb der individuellen Bezugssysteme lagen, ähnlich wie dies oben in Zusammenhang mit den Pfeilwurf-Versuchen schon erörtert wurde.

Es wurde bereits angesprochen, daß in einigen Versuchsserien nicht primär die empfundene und erinnerte Reaktionszeit zu beurteilen war, sondern die jeweils nach der Reaktion vom Versuchsleiter mitgeteilte Zeit. Dies wurde eingeführt, wir deuteten es an, um denjenigen unserer Versuchspersonen, die das Gefühl hatten, allein aufgrund ihres subjektiven Zeitempfindens nicht über längere Serien hin absolut urteilen, sondern nur Paarvergleiche anstellen zu können, über die Zeitwerte eine Hilfe zu geben. Relevant war dies allerdings

insofern nicht, als die Versuchspersonen ja nicht die Dauer des akustischen Signals beurteilen sollten, sondern die Geschwindigkeit ihrer Reaktion auf dieses Signal resp. ihre Zufriedenheit mit dieser Geschwindigkeit - in die freilich die erinnerte Signaldauer mit einging. Tatsächlich jedenfalls läßt die Werteverteilung darauf schließen, daß unsere Versuchspersonen - wie erwartet - absolut urteilten, auch wenn sie selbst das Gefühl hatten, dies nicht zu können. Wir führten also die Mitteilung der Zeitwerte als (subjektive) Hilfestellung ein, zumal uns dies gleichzeitig die Möglichkeit zu bieten schien, über eine "Manipulation" der Zeiten durch gesteuertes feed-back die Bewertungssysteme besser auszuloten, indem wir die Bereichs- und Systemgrenzen eingehender prüften. Dieser zunächst für möglich gehaltene Ausweg auch aus dem Dilemma des oben bereits erörterten schmalen Leistungsspektrums erwies sich jedoch als untauglich, weil praktisch alle Versuchspersonen vor allem in ihren suboptimalen Leistungsbereichen, auf die es hier in erster Linie ankam, Unterschiede von 1cs mit hoher Zuverlässigkeit herausfinden konnten, von größeren "Korrekturen" ganz abgesehen. Meist galt dies für den Bereich von 14-16cs, bei längeren Reaktionszeiten als etwa 19cs ging diese Diskriminationsfähigkeit praktisch vollkommen verloren - was nach dem WEBERschen Gesetz gar nicht anders zu erwarten ist, denn natürlich ist das Verhältnis von 14cs : 15cs ein anderes als das von 19cs : 20cs, natürlich sind die jeweiligen WEBER-Brüche verschieden. Diese Ergebnisse wurden auch in eigens angestellten Kontrollversuchen bestätigt.
Außerdem führte die Bewertung der eigenen Reaktionsleistungen aufgrund der mitgeteilten Zeit-Werte zu relativ starren Beurteilungsstrukturen durch "festklammern" an Zahlen "bis 15 ist gut, ab 20 ist schlecht" o.ä..

Unter diesen Bedingungen der Urteile aufgrund der Zeitwerte

konnten wir auch Drei-Kategorien-Systeme feststellen, allerdings - wie gesagt - ohne auch nur leidlich deutlich sichtbare Dynamik, d.h. über die einzelnen Versuche hin konnten keine nennenswerten Nullpunkt- oder Bereichsverschiebungen beobachtet werden.
Als Beispiel für einen derartigen Versuch mag das in der folgenden Darstellung wiedergegebene Bezugssystem der Vp RE gelten (Darst.4):



Darst.4: Bewertung von Reaktionszeit-Werten im Drei-Kategorien-Versuch: relative Häufigkeiten

Es beruht auf 40 Durchgängen und streut von 10 - 26 cs, der Umfang des Systems beträgt also 17.
Die Zäsuren wären hier nach dem WITTEschen Theorem zu erwarten gewesen bei 13,5cs bzw. 21,5cs - tatsächlich liegen sie bei 13,8cs bzw. 19,9cs. Wir stellen dabei also wiederum - wie schon in den oben berichteten Fällen - gute Übereinstimmung fest, besonders bei den kurzen Reaktionszeiten, im suboptimalen Bereich also, für den die Vpn ein recht gutes System, einen ziemlich zuverlässigen Bewertungsmaßstab ihrer

eigenen Leistung aufgrund sehr feinen Diskriminationsvermögens besitzen, wohingegen bei längeren Zeiten die Einschätzung ein wenig schlechter ist, diese Signallängen sind - wie zu erwarten - schlechter zu diskriminieren.
Insgesamt aber erleben wir auch nach dieser Versuchsanordnung eine Bestätigung unserer Annahmen von der Gleichverteilung der Beurteilungen.

4.4. Zusammenfassung

Die vorstehend berichteten Ergebnisse aus Versuchen zur Beurteilung der eigenen Reaktionszeit (akustisch) im Sinne des Anspruchsniveaus haben trotz mancher Einschränkungen, deren Problematik erst während der Versuchsdurchführungen sichtbar wurde, insgesamt eine Bestätigung unserer Vermutungen gezeigt:
Bezüglich der individuellen Bewertung des eigenen Leistungsverhaltens sind Bezugssystem-Strukturen feststellbar, die denen aus der Beurteilung von Mannigfaltigkeiten, die sich über physikalische Metriken anordnen lassen, zu entsprechen scheinen.
Wenigstens für Bezugssysteme mit zwei und drei Kategorien konnte dargestellt werden, daß die Urteile sich ganz offenkundig - hypothesengerecht - in ihren Kategorien normalverteilen, die Zäsuren zwischen den Kategorialabschnitten liegen - vor allem bei kürzeren Reaktionszeiten - weitgehend exakt dort, wo sie nach WITTE anzunehmen waren.
Eine Einschränkung der Ergebnisse ergibt sich aus dem geringen Umfang der individuell relevanten Leistungsspektren, aus den geringen Mannigfaltigkeiten der zu bewertenden Ereignisse. Diesem Problem soll zur weiteren Absicherung unserer Befunde in der im folgenden Kapitel berichteten Versuchsreihe abgeholfen werden.

5. Weitergehende Analyse von Struktur und Dynamik der individuellen Leistungseinschätzung

5.1. Versuchsplanung

Die im folgenden Kapitel berichtete Versuchsreihe sollte die bisher dargestellten Befunde absichern helfen und darüber hinaus unter Berücksichtigung der angesprochenen versuchstechnischen Schwierigkeiten weitergehende Ergebnisse liefern.

Wir entschieden uns dabei wiederum für das Dart-Spiel, das uns von seiner ganzen Art her für unsere Fragestellung optimal geeignet schien, wie bereits oben erläutert wurde. Die gesamte Leistungssituation dürfte hier - in der für unsere Versuchsabsichten bedeutsamen Form - wesentlich stärker ausgeprägt sein, als sich dies etwa am Reaktionsgerät gezeigt hatte. Zudem zeigen sich beim Dart-Spiel hinreichend viele Möglichkeiten, die in den bisherigen Versuchen gemachten Erfahrungen durch entsprechende Verbesserungen zu nutzen, ohne die Situation grundsätzlich zu verändern:

a. Statt der oben beschriebenen 5-Ring-Scheibe (mit den Werten 10, 25, 50, 75 und 100) benutzen wir jetzt eine 11-Ring-Scheibe (1, 2, ..., 11) gleichen Durchmessers (44cm). Es ergaben sich so Ringe von je 2cm Breite.
Die Scheibe war einheitlich hellgrau, die einzelnen Wertringe lediglich durch dünne schwarze Linien voneinander getrennt. Die Versuchspersonen konnten so auf Wurfentfernung zwar sehen, wie nah der Pfeil am Scheibenmittelpunkt gelandet war, nicht jedoch auch den zugehörigen Wert exakt ablesen; eine relativ starre Zuordnung von Beurteilungsbegriffen zu Ringwerten war erschwert, dies zusätzlich zu der größeren Ring-Anzahl, die zur besseren Ausdifferenzierung der individuellen Leistung und so auch des Beurteilungsverhaltens beitragen sollte.

b. Es wurden je sechs Pfeile gleicher Farbe zur Verfügung gestellt. Die Versuchspersonen hatten zu Beginn jedes Versuches "ihre" Farbe zu wählen (blau, gelb oder rot) und benutzten dann nur diese Pfeile: eine Zuordnung von Beurteilungsbegriffen zu der Farbe des benutzten Pfeils war auf diese Art nicht mehr ohne weiteres möglich, gegebenenfalls konnte sie als Konstante angesehen werden.

c. Die Versuchspersonen bewerten immer und ausschließlich den Summenwert aus je sechs Würfen.
Die Beurteilung einzelner Würfe hätte wiederum - wie bei den Reaktionszeitversuchen - unter einer zu kleinen Mannigfaltigkeit gelitten, die Leistungen und ihre Bewertungen hätten nur in einem Spielraum von 12 "Reizen" streuen können (11 Werte und Fehlversuch). Durch unsere Zusammenfassung war das Spektrum breiter, theoretisch waren als Beurteilungsgrundlage Werte von 0 bis 66 Punkten möglich. Außerdem hätte bei der Einzelwurf-Beurteilung eine zu häufige Bewertungs-"Pflicht" - wie sich ja ebenfalls bereits in unseren früheren Versuchen gezeigt hat - den meisten Vpn Schwierigkeiten gemacht und die Situation sowie damit auch unsere Ergebnisse durch unregelmäßige Stellungnahmen beeinträchtigt. Bei einer Zusammenfassung von jeweils sechs Würfen schien diese Frequenz adäquater.

d. Jede Versuchsperson sollte 50 Serien (entsprechend 300 Würfen) "absolvieren", also dementsprechend auch 50 Bewertungen ihrer eigenen Leistung abgeben.
Diese Zahl schien angebracht, wollten wir angesichts des ja bewußt gewählten theoretisch breiten Leistungsspektrums nicht zu geringe Häufigkeiten bei den einzelnen Werten erhalten. Mehr Wurfserien hätten dagegen aller Erfahrung nach auch bei hochmotivierten Versuchspersonen zu sehr unter dem Faktor [illegible]dung gelitten.

e. Jede Versuchsperson sollte in dieser abschließenden Versuchsreihe an mehreren Versuchen teilnehmen und dabei ihre Leistungen nach 2, 3, ... Kategorien bewerten. Die maximale Kategorien-Anzahl wurde bewußt offen gelassen, da sie ja von der empirischen Werteverteilung und den Differenzierungsmöglichkeiten der verschiedenen Vpn abhing.
Jede Versuchsperson sollte mit niedrigen Kategorien-Anzahlen beginnen und im Zuge wachsender Vertrautheit mit dieser Art der Leistungsmessung - ein Bezugsrahmen für die motorische Leistung im allgemeinen, auch beim Pfeilwerfen, lag ja vor, da das Spiel bekannt war - immer mehr Kategorien verwenden.
Bei dieser Art der Planung folgen wir SIEVERS (im Gegensatz zu den Fraktionierungsversuchen von KELLNER). Da es hier in erster Linie um die Erkundung intraindividueller Bezugssystem-Strukturen geht - von denen aus dann erst auf interindividuelle Verhältnisse geschlossen werden kann - wäre durch für jede Kategorien-Anzahl verschiedene Versuchspersonengruppen (so bei KELLNER) nichts gewonnen gewesen. Vielmehr haben wir so die Möglichkeit, die Leistungsbeurteilungen jeweils aller Vpn unter dem Blickwinkel verschiedener Ausdifferenzierungen zu betrachten - sie, wie etwa SIEVERS es nennt, auf verschiedenen Systemhöhen heranzuziehen.

f. Die Anzahl der Versuchspersonen wurde ebenfalls offen gelassen. Wir setzten aber voraus, daß wir uns nach den bisherigen Erfahrungen auf nur wenige Vpn beschränken konnten, da es bei unserer Arbeit ja nicht primär um eine umfassende Absicherung interindividueller Gemeinsamkeiten geht, sondern zunächst einmal um den Existenznachweis intraindividueller Phänomene auf dem Gebiet der Beurteilung der eigenen Leistung.

5.2. Versuchsdurchführung

Im Sinne der geschilderten Planung wurden insgesamt 19 Versuche durchgeführt. Sie erstreckten sich über mehrere Wochen, zwischen je zwei Versuchen einer Versuchsperson lag mindestens ein Tag Zwischenzeit.
Dabei nahmen die fünf Versuchspersonen dieser abschließenden Versuchsreihe teil an

5 Versuchen nach 2 Kategorien,
6 " " 3 " ,
7 " " 4 " ,
1 " " 5 " .

Jede Versuchsperson bewertete ihre Leistungen mindestens einmal nach 2, 3 und 4 Kategorien; im vierten Versuch wurde jeweils auf dem Hintergrund der bis dahin vorliegenden individuellen Ergebnisse mit den Vpn zusammen die Anzahl der Kategorien festgelegt. Da wir wiederum mit Psychologie-Studenten älterer Semester arbeiteten, schien dies sinnvoll, um möglichst präzise Ergebnisse zu erhalten, ohne die Vpn zu überfordern. Über die engeren Versuchsabsichten waren die Versuchspersonen sich dennoch nicht vollständig im klaren, da wir unsere Versuche als Diskriminationsversuche ausgegeben hatten.

Aufgrund der ausgewählten Stichprobe konnten wir die Versuchsanweisung wiederum frei geben: unsere Vpn sollten

a. einen Summenwert für das Ergebnis aus je sechs Würfen nennen, den sie im Sinne der Anspruchsniveausetzung als für sich "zufriedenstellend" bzw. "gut" werteten;
b. jeden erzielten Summenwert auf dem Hintergrund dieser Anspruchsniveausetzung nach dem Kriterium "zufriedenstellend" bzw. "gut" beurteilen (beide Begriffe wurden von den Vpn als Synonyme verwendet);
c. nach jeder Beurteilung mitteilen, ob und auf welchen Wert das Anspruchsniveau geändert wurde.

Verwendet wurden in den Versuchen die folgenden Einstufungs-Kategorien:

bei 2 Kategorien: gut, schlecht;
bei 3 Kategorien: gut, mittel, schlecht;
bei 4 Kategorien: gut, eher gut, eher schlecht, schlecht;
bei 5 Kategorien: sehr gut, gut, mittel, schlecht, sehr schlecht.

In allen Fällen ist gut/schlecht synonym zu sehen mit zufrieden/nicht zufrieden.

Bei der Wahl der Einstufungs-Begriffe haben wir uns - vor allem bei den Vier-Kategorien-Versuchen - durch die Erfahrungen von KELLNER und SIEVERS leiten lassen; es kann hier vor allem auf die entsprechende Erörterung bei SIEVERS verwiesen werden (S.36ff).

5.3. Ergebnisse

5.3.1. Zum allgemeinen Versuchs- und Beurteilungsverhalten

Alle Versuchspersonen kamen mit der veränderten Versuchsanordnung ohne Probleme zurecht, Änderungen der Planung waren von daher nicht erforderlich. Insgesamt waren dabei im Leistungs- und Beurteilungsverhalten die gleichen grundlegenden Phänomene festzustellen, die sich bereits in den früheren Versuchsreihen angedeutet hatten:

a. Die Farbe der Wurf-Pfeile spielt ganz offensichtlich eine Rolle: Von den 19 Versuchen dieser Versuchsreihe wurden 9 mit roten, 7,5 mit gelben und 2,5 mit blauen Pfeilen geworfen - eine Vp wechselte die Pfeile während des Versuchs (nach einer Mißerfolgsserie!). Teilweise wurden über die verschiedenen Versuche hin die Pfeile konstant gehalten, teilweise systematisch durchpermutiert.

b. Das Entscheidungskriterium wird im Falle des Mißerfolgs

teilweise gewechselt. So wird etwa nicht mehr der Punktwert beurteilt, sondern daß alle Pfeile dicht zusammen sind - und wenn dies auch ganz am Rande der Scheibe ist. Daneben kann der Punktwert von vornherein relativiert werden - in einem Ausmaße, das fast gar keinen Mißerfolg zuläßt, z.B.: "Dafür daß nur vier Pfeile (von sechs) auf der Scheibe sind, ist das Ergebnis doch eigentlich gut!". Tatsächlich wurde das "dafür daß" sogar in noch extremeren Fällen zur Wertung herangezogen - und in den weitaus meisten Fällen war vollkommen klar, daß die Vpn dies durchaus ernst meinten!

c. Es wird teilweise versucht, Mißerfolge zu antizipieren - und sie dadurch sozusagen zu negieren, außerhalb des Gütemaßstabes zu stellen. Ein Beispiel hierfür bietet etwa Vp RR, die meint, jede Wurfserie unmittelbar nach einer Erhöhung des Anspruchsniveau-Wertes werde aufgrund der verstärkten Spannung schlecht. Tritt dieser Fall dann tatsächlich ein, so ist es eine Bestätigung der Vorhersage, der Mißerfolg ist relativiert, ist eigentlich für die Vp gar keiner mehr. (Im anderen Falle - des Erfolges nämlich - wird eine angenehme Überraschung konstatiert; der Erfolg ist - selbstverständlich nach dem oben Gesagten - durchaus ein Erfolg, der mit absoluten Begriffen gewertet wird.)

d. Von Mißerfolgen wird auch sonst teilweise abgelenkt, indem Hypothesen allgemeinerer Art ("Vorher kräftig fluchen, das hilft!") oder auch in ganz spezieller Form aufgestellt und von der Vp "durchgetestet" werden, etwa zur Wurftechnik, zur aufzuwendenden Kraft, zur Flugbahn der Pfeile. Kommt es so (noch immer) zu Mißerfolgen, dann führt dies nicht zur Unzufriedenheit mit der eigenen Leistung im Sinne der Anspruchsniveausetzung, sondern die Hypothese war eben falsch, man hatte die falsche Einstellung.

Wir können damit hier eine Bestätigung der Hinweise finden, die wir im Rahmen der oben dargelgten Kasuistik bereits festgestellt haben. Auch unter den - im Vergleich zu den früheren Versuchsreihen - reduzierten Anregungsbedingungen, bei zweifellos stringenteren Versuchsbedingungen finden unsere Versuchspersonen genügend Mittel und Wege, um ihre Mißerfolgsquoten subjektiv zu verkleinern, indem sie auf andere Beurteilungsmaßstäbe, in andere Bezugssysteme ausweichen. Auch wenn wir hier dennoch die Versuchsbedingungen nicht eben extrem reduziert haben, so können wir doch wohl insgesamt jetzt aufgrund unserer Ergebnisse vermuten, daß auf dem Gebiete des Leistungsverhaltens Atmosphäre und Anordnung der Versuche gar nicht so steril sein können, um den Versuchspersonen nicht doch noch irgendeine Entschuldigung oder vielmehr Erklärung ihres Mißerfolgs zu ermöglichen. Wie wir bereits oben andeuteten, ist dieser Bereich der Person-Umwelt-Bezüge wahrscheinlich zu elementar, als daß jeder Mißerfolg als solcher gewertet werden könnte. Das Anspruchsniveau kann nicht endlos herabgesetzt werden, sondern es gibt hier Systemgrenzen der Gütemaßstäbe, die nicht nur - wie sich in den Ergebnissen unserer ersten Versuchsreihe schon gezeigt hatte - quantitativer Art sein können (z.B. unter 20 Punkten), sondern auch auf der Zeitachse zu lokalisieren sein mögen: ein, zwei oder auch drei Mißerfolge nacheinander mag es geben, aber nicht 10 - hier muß (subjektiv) ein anderer Grund vorliegen, eine andere Erklärung möglich sein. Keine unserer Versuchspersonen ist auf die Idee gekommen (oder hat sich zumindest nichts davon anmerken lassen), daß dergleichen auch an ihr bzw. an ihren zu hoch gesteckten Zielen liegen könnte.
An dieser Stelle mag diese Feststellung genügen, wir werden auf dieses Phänomen jedoch noch näher eingehen im Rahmen der Thematisierung von Anspruchsniveausetzungen und -verschiebungen am Ende dieses Kapitels.

5.3.2. Vorbemerkungen zur Auswertung der Daten

a. Art der Daten

Wir definieren im folgenden die Diskrepanzen zwischen Ist-Werten und Soll-Werten, um die es bei der hier zu behandelnden Auseinandersetzung mit der eigenen Leistung geht, operational als Differenz zwischen mitgeteiltem Zielwert (Anspruchsniveau) und erreichtem Wert (Leistungswert).

Hat eine Vp z.B. "40" als Ziel angegeben und wirft sie 30 Punkte, so beträgt die Differenz "-10". Werden 50 Punkte geworfen, so beträgt (bei gleichem Anspruchsniveau) die Differenz "+10".

Die erreichten Werte werden hier also nicht für sich genommen als Grundlage der individuellen Leistungsbeurteilung gewertet - einerseits sind sie abhängig von der Einteilung der Wurfscheibe, andererseits haben sie interindividuell ganz verschiedene Stellenwerte - sondern, ganz im Sinne unserer Themenstellung, allein in ihrem Bezug zu den jeweiligen Zielwerten.

Für die Differenz von Istwert und Sollwert - statt etwa für ihr Verhältnis im Sinne des WEBERschen Gesetzes - haben wir uns dabei aufgrund der Erfahrungen entschieden, die wir in unseren Versuchen gemacht haben. Bereits in der ersten Versuchsreihe mit dem Dart-Spiel war ja aufgefallen, daß die Versuchspersonen sich sowohl bei der Wahl der Schrittgrößen für ihre Anspruchsniveauverschiebungen (vgl. S.20) als auch teilweise bei den Bereichsfestlegungen (vgl. S.15) maßgeblich an unserem üblichen dekadischen System orientierten: z.B. hatte bei der Veränderung des Summenwertes aus sechs Würfen die Schrittgröße 50 Punkte betragen, gleichgültig ob es sich um den Leistungsbereich um 200 oder - bei derselben Versuchsperson - um 400 Punkte handelte,

gleichgültig, welcher Abschnitt des individuellen Leistungsspektrums gemeint war.
Die Beobachtungen in den Reaktionszeitversuchen hatten diesen Eindruck verstärkt (vgl. S.36) und auch hier in dieser abschließenden Versuchsreihe ist ein entsprechendes Vorgehen der Vpn schlechthin unübersehbar:
Durch die besondere Anfertigung der Wurfscheibe hatten wir versucht, "dekadische Tendenzen" zu durchbrechen und so allzu stereotype Angleichungsversuche zu erschweren. Dies scheint auch gelungen - dennoch zeigen die Daten (die zur Veranschaulichung im Anhang wiedergegeben sind), daß jede Versuchsperson ihre Schrittgrößen zur Anspruchsniveausetzung über die einzelnen Versuche hin konstant und unabhängig vom momentanen Ist- oder Sollwert festlegt - seien es nun im Einzelfall Schritte von 2, 3 oder 5 Punkten. Einzig und allein unmittelbar bei Beginn der Versuche sind teilweise größere Schritte zu vermerken - hier handelt es sich aber wohl lediglich um Grobeinstellungen, die sehr schnell beendet sind. Insofern war unsere Entscheidung für die Differenzenbildung geradezu zwangsläufig. *

b. <u>System- und Kategorialgrenzen</u>
Bei den vorliegenden Versuchen haben wir es mit grundsätzlich anderen Beurteilungsgegenständen zu tun als bei denen mit Gegebenheiten, die sich über physikalische Metriken abbilden lassen, insofern nämlich, als die zu beurteilende Mannigfaltigkeit nicht von vornherein gegeben und nicht für alle Versuchspersonen gleich ist (wie es z.B. mit einer Serie verschieden langer Bleistifte wäre) - vielmehr erstellt jede Vp ihre eigene Mannigfaltigkeit durch ihre Wurfleistungen selbst, praktisch gleichzeitig mit ihren Be-

* Ob wir damit einen probaten Weg gewählt haben, wird noch an weiteren Versuchen zu überprüfen sein. Uns scheint die Konsistenz der vorliegenden Ergebnisse ein positiver Beleg zu sein.

urteilungen. Diese Mannigfaltigkeit kann in der hier gegebenen Versuchsanordnung bei jeweils sechs Würfen im Einzelfall theoretisch 67 Werte annehmen (0-66), das tatsächliche individuelle Leistungsspektrum dürfte hingegen sehr viel schmaler sein, wenn man von wenigen "Ausreißern" absieht.

Da wir nun einerseits bereits oben dargelegt haben, daß auch für die Bewertung der eigenen Leistung individuell Systemgrenzen vorliegen, die zumindest den "extremen Mißerfolg" als Ausreißer qualifizieren (für das Erfolgserleben wurde analoges nicht entsprechend deutlich), da andererseits unsere Versuchspersonen im Eifer des Spiels nicht in jedem Falle deutlich machten, ob sie eine Wurfleistung noch als schlecht oder gut - **oder** schon als Ausreißer, als außerhalb ihres Gütemaßstabes liegend beurteilten, definieren wir für alle Vpn die Systemgrenzen als diejenigen Punkte (Differenzwerte der erläuterten Art), auf die von der Mitte des individuellen Leistungsspektrums einer Vp her gesehen mindestens zwei unbesetzte Ereignisklassen (mögliche Differenzwerte) folgen.

Beispiel:

Vp JH warf im ersten Versuch Summenwerte, deren Differenzen zu den jeweils vorher festgelegten Zielwerten zwischen "+7" und "-26" variierten. Dabei kamen folgende Differenzwerte auch tatsächlich vor:

+7 *	-3	-10	-17 *
+2	-4	-11	-20
+1	-5	-12	-21 *
±0	-6	-13	-24
-1	-7 *	-14 *	-25
-2	-9	-16	-26

Das Differenzen-Spektrum dieser Versuchsperson hat fünf "Lücken" (*), d.h. mögliche Differenzwerte kommen nicht vor: Nach "+7" kommen vier mögliche Differenzen nicht vor, wir definieren daher den System-Anfang mit "+2", dem Wert, auf den eine zunächst ununterbrochene Reihe von Differenzwerten folgt.

Nach "-7" und "-14" ist jeweils nur eine mögliche Klasse von möglichen Differenzen unbesetzt, nach unsererDefinition stellt dies keine für den Systemumfang relevante

Unterbrechung dar.
Nach "-17" sind zwei Ereignisklassen unbesetzt. Da wir ganz offenkundig die Systemmitte von JH bereits überschritten haben, definieren wir "-17" als Systemende. Daß nach "-21" ebenfalls zwei Klassen unbesetzt waren, ist damit bereits ohne Belang.

Analog definieren wir die Bereichsgrenzen der einzelnen Kategorialabschnitte als diejenigen Punkte, auf die von der phänomenalen Mitte eines Kategorialabschnittes eines Beurteilungsmaßstabes aus gesehen (maximale relative Urteilhäufigkeit) mindestens zwei unbesetzte Wertklassen folgen.

Bei diesem Verfahren ergibt die Summe der Kategorialabschnitte nicht notwendig den Systemumfang. Dies hindert aber nicht daran, die Kategorialabschnitte in Bezug auf ihren Umfang zu vergleichen.

Beispiel:
Vp JH verwendet in einem Drei-Kategorien-Versuch die Urteilskategorie "mittel" für die Zielwert-Leistungswert-Differenzen

- 8 zu 50%
-10 zu 100%
-11 zu 75%
-12 zu 25%
-16 zu 33% .

Nach unserer Definition erstreckt sich der Bereich "mittel" von JH in diesem Versuch vom Differenzwert "-8" bis zum Wert "-12".

Dieses Vorgehen erscheint (wie auch bei den Systemgrenzen) angebracht und legitim, kommen doch - um bei dem letzten Beispiel zu bleiben - zwei Differenz-Klassen vor dem Wert "-16" und in allen Klassen nach diesem Betrag ausschließlich "schlecht"-Urteile vor; die eine Bewertung "mittel" bei der Differenz "-16" (bei zwei "schlecht"-Kategorisierungen zum gleichen Wert) kann nur als Zufallsschwankung aufgefaßt werden, die durch die niedrigen Absolut-Häufigkeiten der Bewertungen in den einzelnen Klassen relativ viel Gewicht erhalten hat. Durch unsere operationalen Definitionen streben wir an, eben diese hier kaum vermeidba-

ren Zufallsschwankungen in ihrem Einfluß zu mindern und auf diese Weise zu möglichst zuverlässigen Ergebnissen zu kommen.

c. Verfahren des gleitenden Mittelwertes

Diesem Ziel der möglichst großen Fidelität dient auch das Verfahren des gleitenden Mittelwertes, das wir in den Darstellungen dieses Kapitels anwenden. Wir übernehmen damit ein Verfahren, das etwa bei der Auswertung des PAULI-Tests in der Form des doppelten Ausgleichs angewendet wird. Wir beschränken uns hier jedoch auf den einfachen Ausgleich, weil wir meinen, weitergehende Maßnahmen hätten bei der speziellen Art unserer Daten zu große Verschleifungen zur Folge.

Beispiel:

Vp RR bewertet im Zwei-Kategorien-Versuch alle Wurfleistungen, die zu Differenzen zwischen "+10" und "-1" führen, als "gut". Die weiteren relativen Häufigkeiten der "gut"-Urteile betragen dann

bei -2	100%	bei -6	--
-3	--	-7	33%
-4	50%	-8	--
-5	25%	-9	-- .

In der graphischen Darstellung wird daraus (nach den vorhergehenden 100%-Markierungen) für diesen Teilbereich

bei -3	75%	bei -6	29%
-4,5	37,5%	-7,5	0% *

*: hier also - im Sinne des oben Gesagten - nicht etwa (33 - 0) : 2 % = 16,5%, wie es dem einfachen Ausgleich entspräche, weil auf "-7" zwei unbesetzte Differenzklassen folgen (Bereichsgrenze).

Das gewählte Verfahren führt in einigen Darstellungen zu graphischen Ungenauigkeiten. Wir nehmen dies bewußt in Kauf, weil die Häufigkeitsverläufe so deutlicher werden. Insgesamt haben wir diesem Verfahren der Darstellung, das Zufallsschwankungen relativiert, den Vorrang vor der Alternative der Zusammenfassung von Klassen (Differenzwerten) gegeben. Ein derartiges Vorgehen hätte zwar die Urteilshäufigkeiten in den einzelnen (neugebildeten) Klassen erhöht, durch Verkleinerung der Klassen-Anzahl aber gleichzeitig auch das

individuelle Differenzen-Spektrum eingeschränkt und aus diesem Grunde zu ungenaueren Daten geführt. In diesem Zusammenhang sei auch zurückverwiesen auf die Erfahrungen mit kleineren Ereignis-Mannigfaltigkeiten im Reaktionszeit-Versuch.

5.3.3. Beurteilung der eigenen Leistung nach verschiedenen Anzahlen von Kategorien

Im Sinne der Versuchsplanung wurden - wie bereits angedeutet - mit den Versuchspersonen in dieser Versuchsreihe Experimente zur Beurteilung der eigenen Leistung im Dart-Spiel (auf dem Hintergrund des zuvor jeweils gesetzten Anspruchsniveau-Wertes) durchgeführt, bei denen 2, 3, 4 oder 5 Kategorien vorgegeben waren. Wir erhielten die folgenden Ergebnisse:

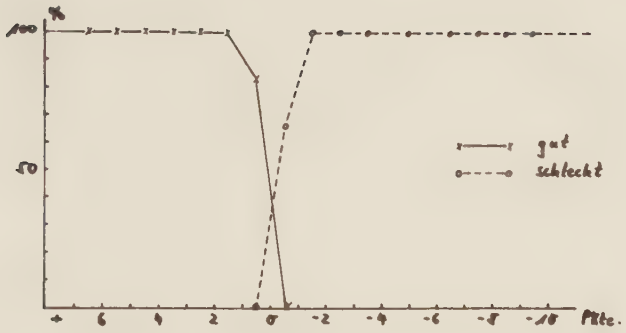
a. Versuche mit 2 Kategorien

Vorgegeben wurden die Begriffe "gut" und "schlecht". Die Ergebnisse sind in der Darstellung 5 (a-e) wiedergegeben, die entsprechenden Kenndaten in der Tabelle 1 zusammengestellt (vgl. Anhang I).
Aus den Daten wird als erstes deutlich, daß die Leistungsbeurteilungen unserer Vpn auf Systemen verschiedener Größe beruhen: die Variationsbreite beträgt hier 18 bis 37 Differenzklassen. Die zugrunde liegenden Leistungen und damit auch deren Diskrepanzen zu den jeweiligen Zielsetzungen haben also ganz unterschiedlich stark gestreut, bei Vp GM (Darst.5c) etwa doppelt so stark wie bei LT (Darst.5a). Dies dürfte in erster Linie mit Unterschieden in der Leistungskonstanz zu erklären sein, beruht darüber hinaus allerdings auch auf der jeweils individuell spezifischen Art der Zielsetzungen. Auf die dabei festzustellenden Unterschiede werden wir noch in einem eigenen Absatz gezielt eingehen.

	Vp	Vers. Nr.	Systemumfang N	von	bis	Ber.Gr. +	−	theor. Mitte	Zäsur erw.	beob.
a.	LT	1	18	+ 7	-10	8	11	- 1,5	- 1,5	± 0,0
b.	RR	1	31	+10	-20	18	17	- 5,0	- 5,0	- 4,2
c.	GM	1	37	+10	-26	17	22	- 8,0	- 8,0	- 5,5
d.	ND	1	20	+ 2	-17	10	13	- 7,5	- 7,5	- 5,0
e.	JH	1	31	+ 2	-28	17	22	-13,0	-13,0	-12,0

Tabelle 1: Kenndaten aus den Zwei-Kategorien-Versuchen

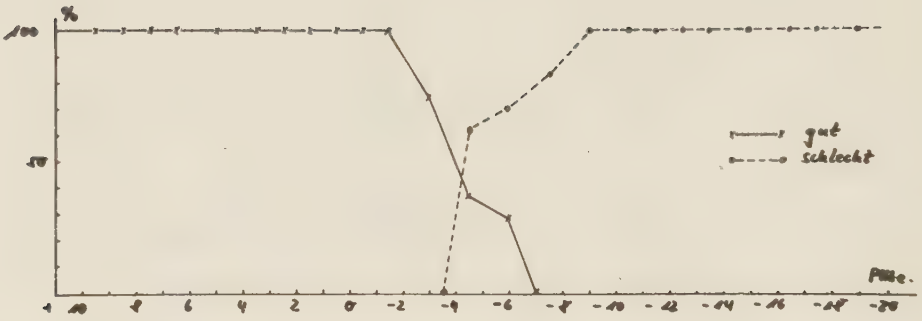
Ber.Gr. = Bereichsgrößen
Bereiche: + = gut
− = schlecht



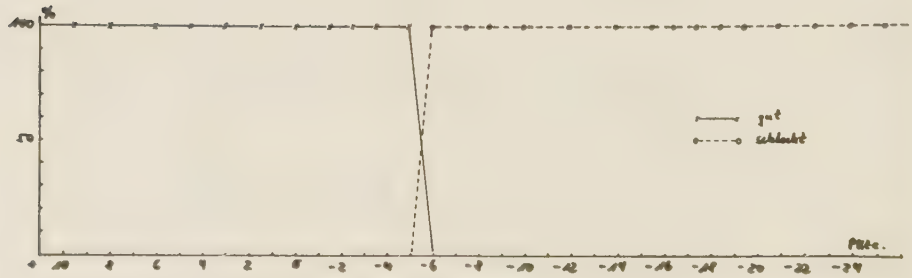
Darst.5a: Differenzen zwischen Ist-Wert und Soll-Wert im Zwei-Kategorien-Versuch: relative Bewertungshäufigkeiten der Vp LT (1.Versuch)

Wir stellen weiterhin fest, daß die System-Mitten bei den verschiedenen Versuchspersonen an ganz verschiedenen Stellen des Differenzen-Spektrums lokalisiert sind; hier reicht die Variationsbreite von -1,5 bis -13,0. Diese Tatsache dürfte ausschließlich aus der Art der Zielsetzungen zu erklären sein und wird deshalb ebenfalls in dem bereits angesprochenen Absatz thematisiert werden.

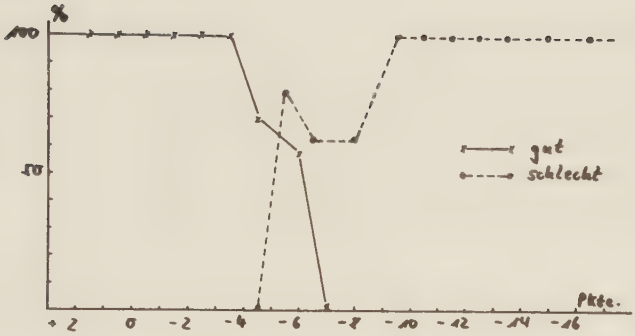
An dieser Stelle kommt es uns in erster Linie auf die Bezugssystem-Charakteristika der Leistungsbeurteilungen an, wie sie sich in den Differenzen-Wertungen ausdrücken. Der Klarheit halber soll dazu hier noch einmal betont werden, daß unsere Vpn die in diesem Kapitel zugrunde gelegten Differenzwerte nicht ausdrücklich beurteilten, sondern implizit: sie nahmen immer zu ihrem jeweils erzielten Leistungswert Stellung. Aus diesem



Darst.5b: Relative Bewertungshäufigkeiten der Vp RR (1.Versuch, 2 Kategorien)

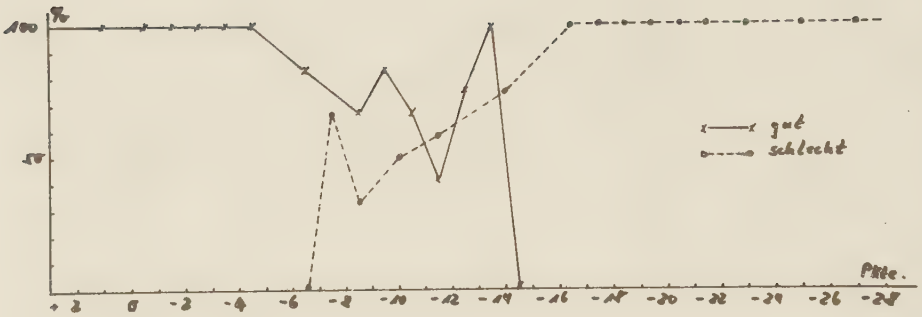


Darst.5c: Relative Bewertungshäufigkeiten der Vp GM (1.Versuch, 2 Kategorien)



Darst.5d: Relative Bewertungshäufigkeiten der Vp ND (1.Versuch, 2 Kategorien)

und dem zuvor mitgeteilten Zielwert im Sinne des Anspruchsniveaus wurde später der Differenzwert berechnet als Operationalisierung der Sollwert-Istwert-Diskrepanz. Bei verschiedenen Wurfleistungen können sich so gleiche Differenzen ergeben - wenn nämlich zwischenzeitlich das Anspruchsniveau entsprechend verschoben wurde.
Wir sehen in den Darstellungen weiter, daß in allen Fällen die Kategorialbereiche wenigstens annähernd gleich groß sind. Genaue Übereinstimmung liegt nur in einem Fall vor (Vp RR, Darst.5b), aber stärkere Ungleichgewichtigkeiten fehlen. In dem - von unseren Hypothesen her gesehen - ungünstigsten Versuch (Vp LT, Darst.5a) überspannt die eine Kategorie ("gut") nur 73% so viele Differenzklassen wie die andere ("schlecht"), bei den anderen Versuchen sind die Relationen zwischen den Kategorialabschnitten ausgeglichener.
Die Lage der Zäsuren läßt sich als noch eher hypothesengerecht bezeichnen: allenfalls in einem Versuch sind größere Abweichungen vom theoretisch zu erwartenden Wert festzustel-



Darst.5e: Relative Bewertungshäufigkeiten der Vp JH (1.Versuch, 2 Kategorien)

len, nämlich bei Vp ND (Darst.5d) mit einer Absolutdifferenz zwischen erwarteter und beobachteter Zäsur von 2,5 Punkten, was auf ihren individuellen Systemumfang (20) bezogen eine Diskrepanz von 13% bedeutet. In den übrigen Versuchen betragen die entsprechenden Beträge (in der Reihenfolge der Darstellungen) 8%, 3%, 7% und 3%.

Bemerkenswert ist hier allerdings zusätzlich eines: Wenn wir sagten, die Lage der Zäsuren sei insgesamt als hypothesengerecht zu bezeichnen, so ist noch modifizierend darauf hinzuweisen, daß (a) praktisch alle Werte negative Differenzen beinhalten und daß dabei (b) die erwarteten Werte "noch negativer" sind als die beobachteten. Es liegen also nicht nur die Nullpunkte der Gütemaßstäbe durchgängig unter den jeweiligen Anspruchsniveaus, sondern es wird auch gleichzeitig deutlich, daß die Vpn hinsichtlich ihrer Wurfleistungen offenkundig mehr erwarten als sie dann einlösen - ein im Sinne unserer Themenstellung wertvoller Befund, der im Zusammenhang mit den entsprechenden Ergebnissen aus den übrigen Ver-

suchen noch in einem eigenen Absatz (5.3.4.) zu thematisieren sein wird.
Vom graphischen Verlauf her fällt vor allem die Darstellung 5c auf (Vp GM), die auf absolut klare Bereichsgrenzen schließen läßt. Tatsächlich hat diese Versuchsperson - bezogen auf ihre verschiedenen jeweils gesetzten Zielwerte - ihre Wurfleistungen derart beurteilt, daß absolut alle im Einzelfall sich ergebenden Differenzen (implizit!) nur mit jeweils einer Kategorie ("gut" oder "schlecht") beurteilt wurden - selbstverständlich ohne daß sie selbst vom rechnerischen Wert dieser Differenz wußte. GM beurteilt auf diese Weise alle Differenzen von +10 bis -4 und dann wieder -6 als "gut", alle von -5 und dann von -7 bis -26 demgegenüber als "schlecht" im Sinne ihres Anspruchsniveaus; ihr Unsicherheitsbereich zwischen den Kategorialabschnitten ist somit extrem schmal, er umfaßt imgrunde nur 2 der in diesem Versuch relevanten 37 Werteklassen.
Umgekehrt läßt z.B. die Darstellung 5e, die die Daten und Urteile der Vp JH wiedergibt, auf einen erheblich größeren Unsicherheitsbereich schließen, es sind sogar drei Zäsuren vorhanden. Die Mitte dieses Zwischenbereichs liegt dennoch mit einer Abweichung von nur 3% vom erwarteten Wert hypothesengerecht - die starken Unregelmäßigkeiten erklären wir demzufolge primär mit den niedrigen Werthäufigkeiten in den einzelnen Klassen, die hier - trotz des Ausgleichs der Daten - noch zu starken Zufallsschwankungen führen. Wir glauben mit gutem Grund annehmen zu können, daß es bei einer längeren Wurfserie (die aus motorischen Gründen letztlich nicht möglich war) zu einer weitergehenden Stabilisierung gekommen wäre. Ein weiterer Grund für unsere Annahme und Interpretation ist das Zielsetzungsverhalten speziell dieser Versuchsperson, das sich u.a. etwa in der Systemmitte von -13 ausdrückt; dieser Punkt soll jedoch am Ende dieses Kapitels noch eigens behandelt werden.

b. Versuche mit 3 Kategorien

Vorgegeben wurden die Begriffe "gut", "mittel", "schlecht". Die Ergebnisse sind in der Darstellung 6 (a-f) wiedergegeben, die entsprechenden Kenndaten in der Tabelle 2 zusammengestellt (vgl. Anhang II).

Wir sehen auch hier wieder verschiedene Systemumfänge bei den einzelnen Versuchspersonen hinsichtlich ihrer Leistungsbeurteilungen; allerdings sind die interindividuellen Unterschiede deutlich nicht mehr so groß wie bei den Zwei-Kategorien-Versuchen, die Variationsbreite erstreckt sich lediglich noch von 17 bis 28 Differenzen-Klassen. Offenkundig hat eine Stabilisierung der Leistung, eine Verringerung der Streubreite stattgefunden; der für die Beurteilung relevante Bereich hat sich dadurch verkleinert.

Daneben mag es auch zu einer Nuancierung bei den Zielsetzungen gekommen sein, wodurch große Differenzen - vor allem mit negativem Vorzeichen - zwischen Ist-Wert und Soll-Wert entfallen würden; einschlägig ist nach der Datenanalyse in jedem Falle primär der erste Grund, zumal auch die Unterschiede zwischen erwarteten und beobachteten Zäsuren nicht mehr gleichsinnig sind.

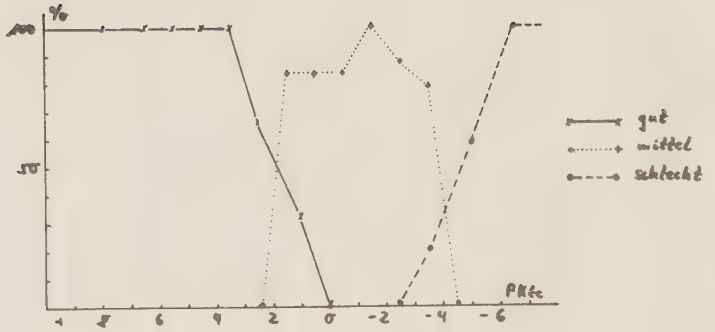
Was das interindividuell unterschiedliche Zielsetzungsverhalten generell betrifft, das sich u.a. auch wiederum in den verschiedenen System-Mitten ausdrückt, so verweisen wir dazu auch an dieser Stelle auf den entsprechenden Absatz am Ende dieses Kapitels, in dem dieser Aspekt für sich thematisiert wird.

Bezüglich der Bezugssystem-Eigenschaften des Bewertungs-Verhaltens unserer Versuchspersonen sehen wir hier schon deutlichere Unterschiede, als dies bei den Zwei-Kategorien-Versuchen der Fall war. Auf der einen Seite liegen Ergebnisse vor, bei denen schon die graphische Darstellung den Eindruck einer hypothesenentsprechenden Verteilung der Urteile nahelegt - etwa die Darstellungen 6a (Vp LT) und 6c (Vp GM);

	Vp	Vers. Nr.	Systemumfang N	von	bis	Bereichsgröß. +	0	-	theor. Mitte	Zäsuren erw.	beob.
a.	LT	2	17	+9	- 7	10	7	5	+ 1,0	+ 2,7 - 2,7	+ 1,9 - 4,2
b.	RR	2	23	+3	-19	9	14	11	- 8,0	- 3,7 -11,3	- 4,8 -11,0
c.	GM	2	28	+9	-18	11	8	12	- 4,5	+ 0,7 - 8,7	- 0,5 - 7,0
d.	ND	2	23	+7	-15	11	2	11	- 4,0	+ 0,3 - 7,3	- 3,0 - 4,8
e.	JH	2	20	±0	-19	13	5	8	- 9,5	- 5,7 -12,3	- 8,9 -12,0
f.	JH	4	24	-5	-28	6	8	12	-16,5	-12,0 -20,0	- 9,3 -16,5

Tabelle 2: Kenndaten aus den Drei-Kategorien-Versuchen

Bereiche: + = gut
0 = mittel
- = schlecht

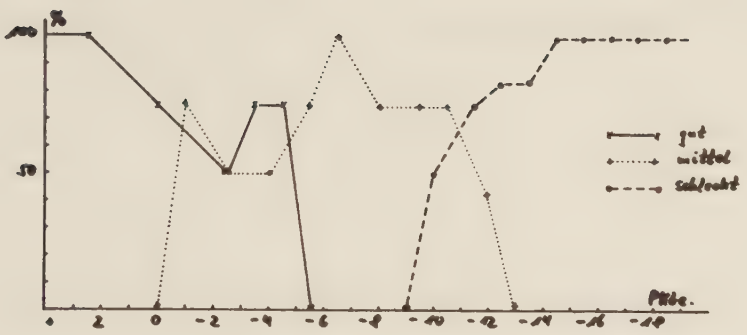


Darst.6a: Differenzen zwischen Ist-Wert und Soll-Wert im Drei-Kategorien-Versuch: relative Bewertungshäufigkeiten der Vp LT (2.Versuch)

andererseits stehen ebenso Befunde dagegen, die zwar in der Leistungsbewertung drei Kategorien aufweisen, jedoch vermuten lassen, daß eigentlich nur zwei Kategorialabschnitte realisiert sind. Beispielhaft dafür sind vor allem die Urteilsverteilungen der Vp ND (Darst.6d) und der Vp JH (Darstellung 6e).

Die Kennwerte bestätigen diesen Eindruck. In den beiden erstgenannten Fällen sowie außerdem im Versuch mit der Vp RR (Darst.6b) sind die Diskrepanzen zwischen erwarteten und beobachteten Kategorialzäsuren klein, teilweise sogar sehr klein; sie erreichen hier - bezogen auf den jeweiligen System-Umfang - in den insgesamt sechs Einzelfällen (zwei Zäsuren je Versuch) Werte zwischen 1% und 9%.

Anders in den Versuchen d (Vp ND), e und f (beide Vp JH). Die Zäsuren-Abweichungen sind hier - zum Teil erheblich - größer, sie sind durch Werte zwischen 11% und 17% gekennzeichnet. Aufgrund der Veranschaulichung in den graphischen Darstellungen war bereits zu vermuten gewesen, daß es sich



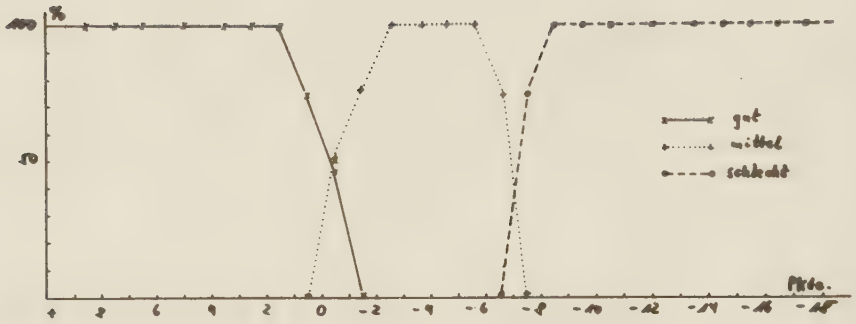
Darst.6b: Relative Bewertungshäufigkeiten der Vp RR (2.Versuch, 3 Kategorien)

hier wenigstens teilweise eigentlich um Zwei-Kategorien-Ergebnisse handelt, auch wenn die Versuchspersonen verbal drei Kategorien zur Bewertung ihrer Leistung verwenden. Dergleichen würde sich decken mit Beobachtungen, die bereits seinerzeit in Tübingen von WITTE und seinen Mitarbeitern gemacht wurden. Sie beinhalten, daß nachfolgende Kategorisierung mit mehr Kategorien als zuvor - wie hier in unserem Fall - in der Regel zu kleineren Mittelbereichen führt; offenbar wirkt hier die vorangegangene Beurteilung nach.
Dieser Eindruck bestätigt sich rechnerisch für alle drei genannten Versuche: Wir fassen dazu - vor allem nach Maßgabe der (aus Tabelle 2 ersichtlichen) Bereichsumfänge - je zwei Kategorialabschnitte zusammen und zwar

(1) für den Versuch d "gut" und "mittel" gegenüber "schlecht";

(2) für den Versuch e "mittel" und "schlecht" gegenüber "gut" und

(3) für den Versuch f "gut" und "mittel" gegenüber "schlecht".



Darst.6c: Relstive Bewertungshäufigkeiten der Vp GM (2.Versuch, 3 Kategorien)

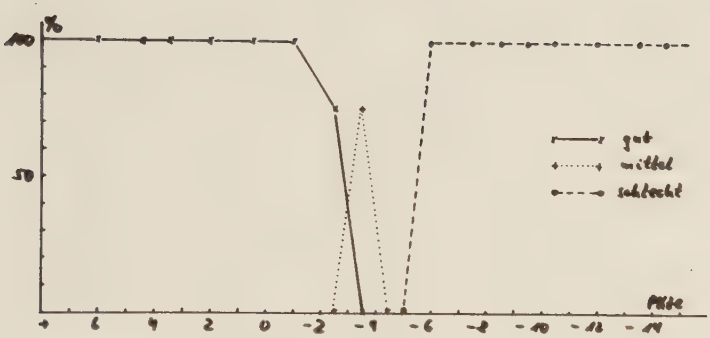
Die damit notwendigerweise neu zu erwartenden bzw. festzustellenden Zäsuren betragen für den Versuch

d. erwartet: - 4,0 und beobachtet: - 4,8;
e. " - 9,5 " - 8,9;
f. " -16,5 " -16,5.

Bereits auf den ersten Blick zeigt der Vergleich hohe Übereinstimmung, tatsächlich betragen die Abweichungen der beobachteten Zäsuren von den erwarteten maximal 3% - bezogen wiederum auf die jeweiligen Systemumfänge unserer Versuchspersonen.

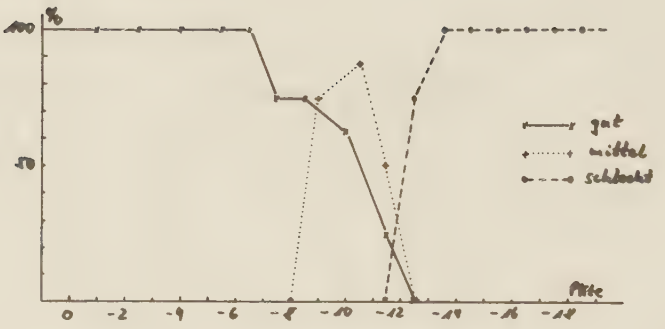
Wir müssen also annehmen, daß diese Versuchspersonen für die Bewertung ihrer Pfeilwurf-Leistungen kein derart ausdifferenziertes Bezugssystem, keinen derart feinen Gütemaßstab besitzen, daß ihnen eine Beurteilung nach drei Kategorien möglich ist; sie können tatsächlich wohl nur nach "gut" und "schlecht" unterscheiden, auch wenn sie verbal feiner zu differenzieren versuchen.

Man könnte zunächst vermuten, daß dies hier prinzipiell un-

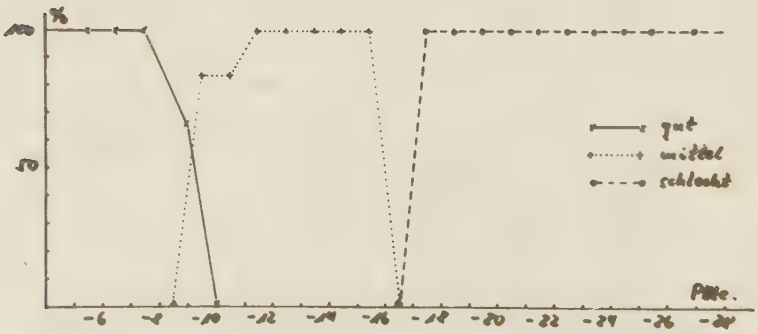


Darst.6d: Relative Bewertungshäufigkeiten der Vp ND (2.Versuch, 3 Kategorien)

möglich sei. Denkbar wäre ja durchaus, daß es ein besonderes Charakteristikum der Ereignismannigfaltigkeit "Leistung" - oder wenigstens "Leistung im Pfeilwerfen" - ist, wenn es kognitiv nur zwei Kategorialbereiche gibt: zufrieden oder nicht zufrieden, gut oder schlecht, Ziel erreicht oder Ziel nicht erreicht - die normalerweise elementare Bedeutsamkeit dieses Person-Umwelt-Bezuges sensu HECKHAUSEN ließe dies durchaus plausibel erscheinen.
Diese Hypothese läßt sich jedoch angesichts der übrigen Befunde nicht aufrecht erhalten; feinere Differenzierungen bei der Beurteilung der eigenen Leistung sind möglich und werden in unseren Versuchen auch vorgenommen, wenigstens von einigen Vpn.
Diesen Eindruck bestätigt auch das bislang hier noch unerwähnt gebliebene Versuchs-Ergebnis der Vp RR (Darst.6b): Trotz einiger Unsicherheiten, die vor allem in der graphischen Darstellung bei der Unterscheidung von "gut" und "mittel" zum Ausdruck kommen, läßt sich feststellen, daß RR -



Darst.6e: Relative Bewertungshäufigkeiten der Vp JH (2.Versuch, 3 Kategorien)



Darst.6f: Relative Bewertungshäufigkeiten der Vp JH (4.Versuch, 3 Kategorien)

wie auch LT und GM - die Bewertung seiner Pfeilwurf-Leistungen in drei Kategorien differenziert. Die Zäsuren zwischen den Kategorialabschnitten liegen mit Abweichungen von nur 5% (gut-mittel) bzw. von 1% (mittel-schlecht) von den erwarteten Werten hypothesengerecht - die Umfänge der Kategorialabschnitte sind tatsächlich etwa gleich.
Wir stellen somit bei drei Versuchspersonen bezüglich der Beurteilung der eigenen Leistung im Drei-Kategorien-Versuch eine insgesamt erwartungsgemäße Verteilung der Wertungen fest; drei andere Versuche (von zwei Vpn) lassen trotz der Verwendung von drei verschiedenen Begriffs-Stufen vermuten, daß die Versuchspersonen ihre Wurfleistung kognitiv eigentlich in nur zwei Bereiche strukturiert haben; die dritte Wertung kommt bei ihnen zustande, indem sie einen Abschnitt verbal fraktionieren (wohl ohne daß ihnen dies recht klar ist!): in diesen Fällen verhalten sich die Umfänge der Kategorialabschnitte praktisch nicht wie 1:1:1 (wie nach unseren Hypothesen zu erwarten gewesen wäre), sondern eher wie 2:1:1 oder wie 1:1:2.

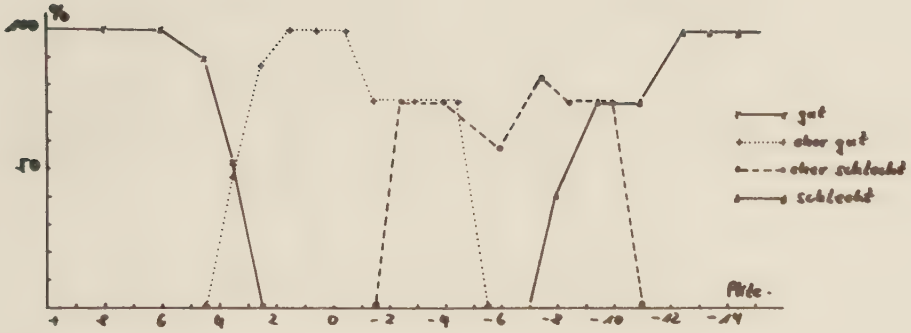
c. Versuche mit 4 Kategorien

Vorgegeben wurden die Begriffe "gut", "eher gut", "eher schlecht", "schlecht".
Die Ergebnisse sind in der Darstellung 7 (a-g) wiedergegeben, die entsprechenden Kenndaten in der Tabelle 3 zusammengestellt (vgl. Anhang III).
Was sich bereits bei den Drei-Kategorien-Versuchen andeutete, wird hier noch deutlicher: einige Versuchspersonen verfügen über einen so fein ausdifferenzierten Gütemaßstab zur Bewertung ihrer Wurfleistungen, daß sie auch imstande sind, diese wirklich nach vier Kategorien zu beurteilen - anderen Versuchspersonen hingegen scheint dergleichen nicht in dieser Form möglich gewesen zu sein.
Vor allen Dingen aber handelt es sich hier jeweils um die

	Vp	Vers. Nr.	Systemumfang N	von	bis	Bereichsgrößen +	:	.	-	theor. Mitte	Zäsuren erw.	beob.
a.	LT	3	25	+9	-15	7	10	10	9	- 3,0	+ 3,3 - 3,5 - 9,8	+ 3,5 - 3,2 - 9,9
b.	RR	3	21	+4	-16	5	10	8	7	- 6,0	- 0,3 - 5,5 -10,8	± 0,0 - 5,5 -10,0
c.	RR	4	21	+9	-11	11	11	5	3	- 0,5	+ 5,0 ± 0,0 - 5,0	- 0,5 - 6,7 - 9,6
d.	GM	3	26	+6	-19	6	10	3	15	- 6,5	+ 0,5 - 6,0 -12,5	+ 1,2 - 5,5 - 7,0
e.	GM	4	27	+9	-17	11	3	5	16	- 4,0	+ 3,3 - 3,5 -10,3	- 1,0 - 3,5 - 7,3
f.	ND	3	26	+2	-23	8	4	7	9	-10,5	- 3,5 -10,0 -16,5	- 6,3 -10,0 -14,5
g.	JH	3	19	-3	-21	4	10	7	6	-12,0	- 6,8 -11,5 -16,3	- 6,8 -15,0 -18,8

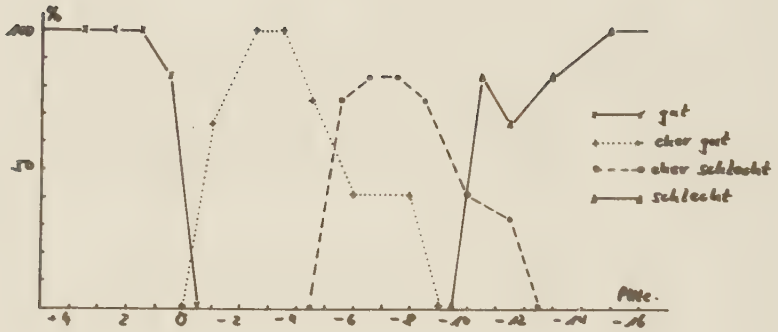
Tabelle 3: Kenndaten aus den Vier-Kategorien-Versuchen

Bereiche: + = gut
: = eher gut
. = eher schlecht
- = schlecht

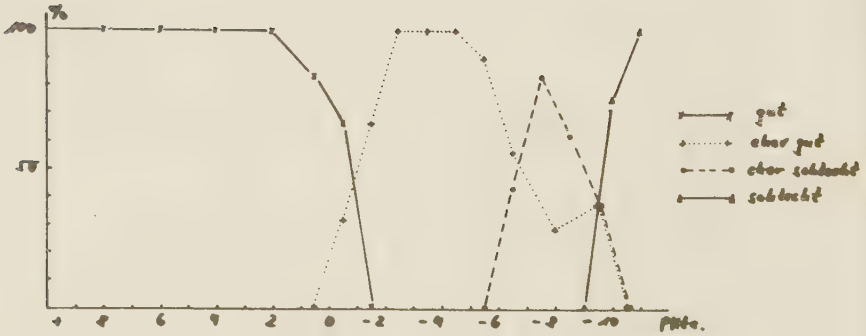


Darst.7a: Differenzen zwischen Ist-Wert und Soll-Wert im Vier-Kategorien-Versuch: relative Bewertungshäufigkeiten der Vp LT (3.Versuch)

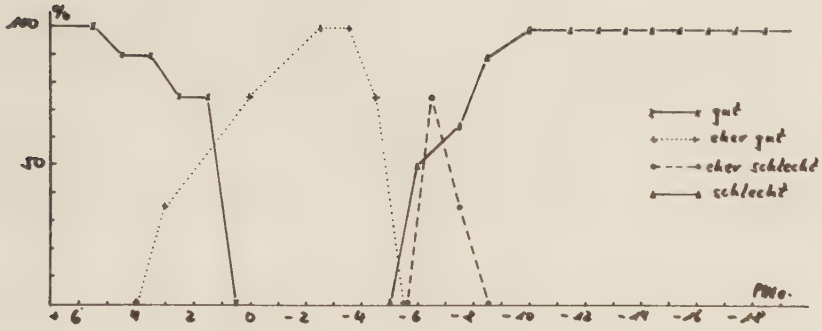
selben "Untergruppen" von Vpn, von denen schon im letzten Absatz die Rede war: In erster Linie LT (Darst.7a), daneben auch RR (Darst.7b), hat mit dieser Art der Versuchsdurchführung kaum Probleme. Bei LT liegen selbst die Zäsuren mit Abweichungen von maximal 1% an den erwarteten Stellen des Differenzen-Spektrums; tatsächlich sind die Kategorialabschnitte insgesamt etwa gleich umfangreich, die Randbereiche dabei ein wenig kleiner als die in der Mitte.
Vp RR kann hier zunächst mit einer Abweichung von höchstens 4% (Zäsur "eher gut" - "gut") ähnlich eingestuft werden, erbringt aber dann bei einer Wiederholung dieser Variante in ihrem vierten Versuch (Darst.7c) erheblich "schlechtere" Ergebnisse im Sinne unserer Vermutungen. In diesem letzten Fall lassen ihre Urteile eher auf eine kognitive Drei-Teilung des persönlichen Wertungsmaßstabes schließen, die Kategorial-Zäsuren liegen für einen Vier-Kategorien-Versuch auch nicht annähernd an den erwarteten Stellen (vgl. auch Tabelle 3).



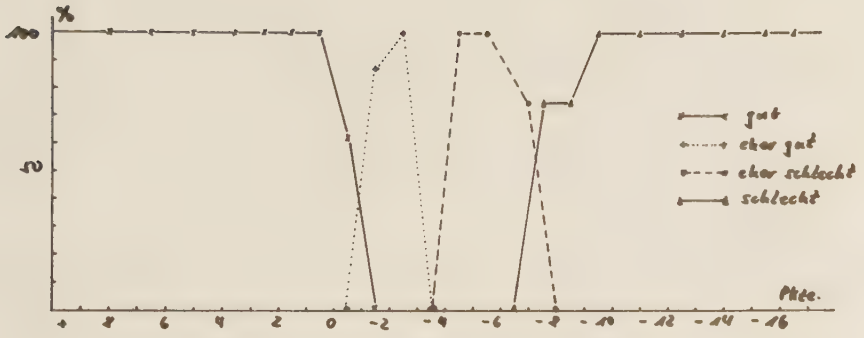
Darst.7b: Relative Bewertungshäufigkeiten der Vp RR (3.Versuch, 4 Kategorien)



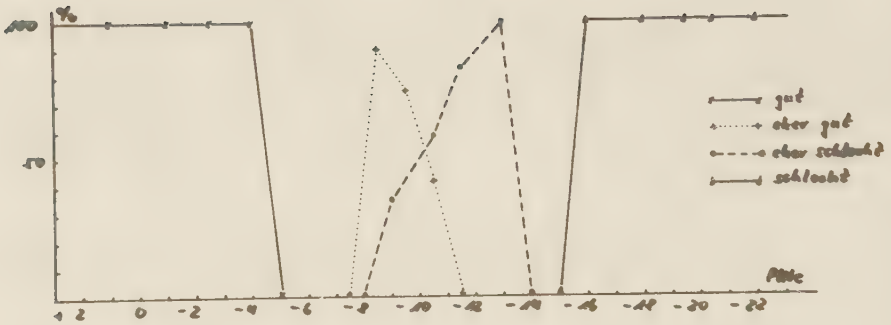
Darst.7c: Relative Bewertungshäufigkeiten der Vp RR (4.Versuch, 4 Kategorien)



Darst.7d: Relative Bewertungshäufigkeiten der Vp GM (3.Versuch, 4 Kategorien)



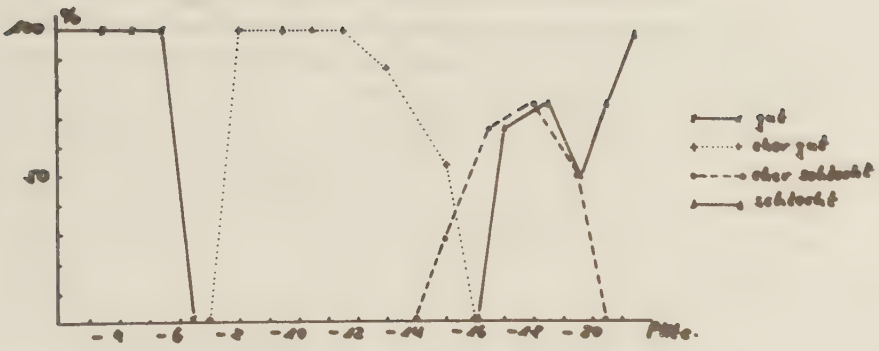
Darst.7e: Relative Bewertungshäufigkeiten der Vp GM (4.Versuch, 4 Kategorien)



Darst.7f: Relative Bewertungshäufigkeiten der Vp ND (3.Versuch, 4 Kategorien)

Alle anderen Versuchsergebnisse nehmen hier eine Zwischenstellung ein, klare Vier-Kategorien-Befunde zeigen sich dabei in keinem Falle. Vielmehr läßt die Verteilung der Daten teilweise auf eine Drei-Teilung (Vp GM im vierten Versuch, Darst. 7e; Vp ND, Darst. 7f) schließen, teilweise auf eine Zwei-Teilung (Vp GM im dritten Versuch, Darst.7d; Vp JH, Darst.7g). Die aus den graphischen Darstellungen bereits ablesbaren Tendenzen bezüglich der Strukturen der individuellen Gütemaßstäbe werden durch die Kenndaten bestätigt und präzisiert.

Bei der Vp GM ist das Versuchsergebnis - ähnlich wie bei RR - nicht eindeutig. Sie nahm ebenso wie diese an zwei Vier-Kategorien-Versuchen teil und diskriminierte dabei verschieden stark. Im ersten Fall (Darst.7d) liegt ganz offenkundig eigentlich eine nur zwei-kategoriale Urteilsverteilung vor, besonders die Kategorie "eher schlecht" deckt praktisch lediglich einen Unterbereich von "schlecht" ab. Die entsprechende wertmäßige Kennzeichnung verdeutlicht, daß in diesem



Darst.7g: Relative Bewertungshäufigkeiten der Vp JH (3.Versuch, 4 Kategorien)

Falle bei entsprechender Zusammenfassung der beiden gut-Kategorien gegenüber dem gesamten schlecht-Bereich die Abweichung der beobachteten Zäsur von der erwarteten nur 2% beträgt. Die Wiederholung dieser Versuchsvariante bei GM (Darst.7e) legt dagegen eher die Vermutung auf eine dreikategoriale Strukturierung des Wertungsmaßstabes nahe, nämlich durch Zusammenfassung der beiden mittleren Bereiche. Dieser gesamte Mittelbereich ist auch dann allerdings noch in seinem Umfange auf dem Differenzen-Spektrum von GM kleiner, als er nach unseren Hypothesen sein sollte. Hierdurch kommen auch die nicht ganz befriedigenden Zäsur-Abweichungen von 7% bzw. 5% - bezogen auf dieses individuelle Spektrum - zustande.

Bei der Vp ND (Darst.7f) haben wir es demgegenüber ganz klar mit einer Ausdifferenzierung der Beurteilung der eigenen Wurfleistung in eigentlich drei Bereiche zu tun: ND differenziert zwar in "eher gut" und "eher schlecht" - tatsäch-

lich liegen beide Bereiche jedoch praktisch übereinander, gibt es demzufolge imgrunde nur einen Mittelbereich. Wird dieser Eindruck schon durch die graphische Darstellung der Befunde dem Beobachter aufgedrängt, so wird er durch die Kenndaten deutlich bestätigt: die (drei) Bereiche sind etwa gleich groß; die neu zu berechnenden bzw. festzustellenden Zäsuren weisen lediglich Diskrepanzen von 2% (gut-"mittel") bzw. 0% ("mittel"-schlecht) zwischen erwarteten und beobachteten Werten auf; trotz der Vorgabe, trotz auch der Verwendung von vier Kategorien - die Ergebnisse dokumentieren für diesen Versuch, daß ND für die Bewertung seiner Wurfleistungen offenkundig nur drei verschiedene Qualitätsstufen kennt.

Ganz anders JH: Bei ihm (vgl. Darst.7g) fallen einerseits die Bereiche "eher schlecht" und "schlecht" kognitiv praktisch zusammen, auf der anderen Seite ist der Umfang der Kategorie "eher gut" so groß, daß daneben nur noch relativ wenig Raum auf dem Urteilsspektrum ist. Am ehesten lassen die Befunde hier auf eine in Wirklichkeit zwei-kategoriale Verteilung der Wertungen mit deutlich verschobener Zäsur schliessen. Dies würde vor allem hinsichtlich der Verschiebung auch in Einklang stehen mit der besonderen Nullpunkt-Lage dieses Gütemaßstabes, nämlich bei einer schon recht extrem scheinenden Istwert-Sollwert-Diskrepanz von 12 Negativ-Punkten, von der unten noch zu sprechen sein wird. Für die Zwei-Zahl der Bereiche spräche daneben bestätigend auch das Ergebnis von JH aus seinen beiden Drei-Kategorien-Versuchen.

Wir sehen demnach auch hier in den Vier-Kategorien-Versuchen, daß es auf der einen Seite Vpn gibt, deren Bewertungsmaßstab für die eigene Wurfleistung soweit ausdifferenziert ist, daß sie nach vier Kategorien urteilen können - und dies nicht nur verbal, sondern in der Form, daß die einzelnen Kategorialabschnitte gleich groß sind (wie wir in unseren Hypothesen vermutet hatten), daß dem also wahrscheinlich auch eine

entsprechende kognitive Strukturierung zugrunde liegt. Auf der anderen Seite müssen wir feststellen, daß dies nur für die Minderheit der in unseren Versuchen mitarbeitenden Vpn gilt; die Mehrzahl scheint kognitiv eigentlich nur zwei Bereiche - "gut" und "schlecht" - strukturiert zu haben.

d. Versuch mit 5 Kategorien

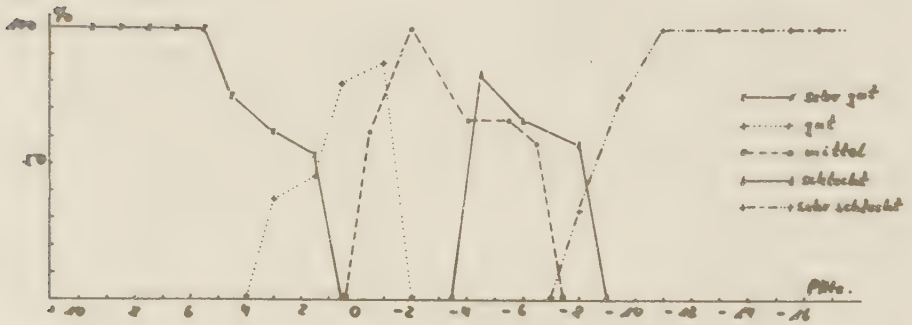
Die anfangs durchaus für realistisch gehaltenen Fünf-Kategorien-Versuche, für deren Möglichkeit ja schon in unserer ersten Versuchsreihe Hinweise gefunden wurden (vgl. Darst.1, S.16) wurden nach den bisherigen Erfahrungen dieser letzten, abschließenden Reihe aus der Planung gestrichen.

Wir wollen hier die in unseren Vier-Kategorien-Versuchen zutage getretene Problematik nicht unbedingt extrapolieren, da wir einerseits die gerade bei vier-kategorialer Beurteilung immer wieder festzustellenden Schwierigkeiten kennen und da wir andererseits - darüber hinaus - für nicht sehr wahrscheinlich halten, daß in jedem Falle vier Bereiche das kognitive Zwischenstadium zwischen drei- und fünfkategorialer Differenzierung und Beurteilung darstellen müssen - WITTE wies hierauf schon 1960 hin.

Bei einem abschließenden Versuch äußerte aber dann die Vp LT einen entsprechenden Wunsch. Da LT in allen bisherigen Versuchen gezeigt hatte, daß sie einen offenbar sehr fein ausdifferenzierten Maßstab zur Bewertung der eigenen Wurfleistungen besitzt, erwarteten wir, daß sich dies auch im Fünf-Kategorien-Versuch spiegeln würde. Wenigstens in diesem einen Falle wurde dann noch die fünf-kategoriale Beurteilung angegangen; die Ergebnisse sind in der Darstellung 8 sowie in den an diese anschließend festgehaltenen Kenndaten zusammengestellt (vgl. Anhang IV).

Bereits die Darstellung veranschaulicht ein Gut-Teil Bezugssystem-Genese: Tatsächlich gliedern sich unübersehbar aus einer ganz deutlichen Drei-Gliederung (!) des Wertungsmaß-

stabes zwei Zwischenbereiche heraus. Die Kenndaten - Zäsuren und Bereichsumfänge - belegen dies. Interessant ist dabei vor allem, daß sich die beiden Zwischenbereiche "gut" und "schlecht" nicht so sehr von den Extremkategorien "sehr gut" und "sehr schlecht" absulösen scheinen, sondern vom Mittelbereich. Zwischen den Extrempolen scheint es zunächst einen mittleren Abschnitt zu geben, der sich dann ausdifferenziert.



Darst.8: Differenzen zwischen Ist-Wert und Soll-Wert im Fünf-Kategorien-Versuch: relative Bewertungshäufigkeiten der Vp LT (4.Versuch)

Kenndaten dieses Versuchs:

Vp LT, Versuch Nr. 4
Systemumfang N = 28, von +10 bis -17
Bereichsgrößen: sehr gut 10
gut 7
mittel 8
schlecht 6
sehr schlecht 11
theoretische Mitte: -3,5
Zäsuren: erw. + 5,4 beob. +1,5
- 0,2 -1,1
- 5,8 -4,2
-11,4 -8,3

Daß dabei die beiden hier sogenannten Zwischenkategorien nicht etwa Unterbereiche des mittleren Abschnittes sind, wird durch die Bereichsumfänge klar: "sehr gut"- und "gut"-Urteile umspannen ebenso viele Differenz-Klassen (17) wie "sehr schlecht" und "schlecht", während die Kategorie "mittel" innerhalb des Leistungsspektrums nur halb so viele (8) Werte umfaßt. Dennoch liegen die beiden zentralen Zäsuren (gut-mittel und mittel-schlecht) deutlich näher bei den erwarteten Werten als die peripheren (sehr gut - gut und schlecht - sehr schlecht), dies aufgrund der bereits angesprochenen phänomenalen Differenzierung des zugrunde liegenden Bezugssystems von der Mitte aus.
Es scheint demnach bei dieser Versuchsperson mit ihrem sehr fein durchstrukturierten Wertungsmaßstab so zu sein, daß eine Wertung nach fünf Kategorien möglich ist. Die Verteilung der Beurteilungen scheint dabei in die Richtung zu gehen, die wir in unseren Hypothesen vermuteten. Eine bestätigende Wiederholung dieses Versuchs war leider aus äußeren Gründen nicht möglich, analoge Ergebnisse von anderen Vpn liegen - wie bereits angesprochen - in dieser Form nicht vor. Mögliche Gründe dafür werden im Zuge der zusammenfassenden Erörterung zu diskutieren sein. An dieser Stelle mag die Feststellung genügen, daß es offenbar möglich ist, auch im Fünf-Kategorien-Versuch hinsichtlich der Beurteilung der eigenen Leistung im Pfeilwerfen Befunde zu erhalten, die so verteilt sind, wie sie nach den WITTEschen Annahmen verteilt sein sollten. Wie groß der Prozentsatz der Versuchspersonen ist, bei denen man derart fein ausdifferenzierte Wertmaßstäbe findet, sehen wir demgegenüber als nachgeordnete Frage an.

5.3.4. Zur Lage der individuellen System-Nullpunkte

Bei der bisherigen Darlegung unserer Versuchsergebnisse haben wir aus Gründen der Übersichtlichkeit die individuellen Urteils-

verteilungen so behandelt, als hätten wir es mit einem konstanten Nullpunkt, mit interindividuell gleichen System-Mitten zu tun. Daß dies nicht der Fall ist, war dennoch unübersehbar; wir wollen diesen Aspekt daher nun hier für sich genommen thematisieren. Zur Verdeutlichung verwenden wir wieder die Daten der letzten Versuchsreihe, die einerseits quantitativ am ehesten interindividuell vergleichbar sind, andererseits in den vorangegangenen Abschnitten unter anderen Gesichtspunkten schon am eingehendsten geschildert wurden.

Als Nullpunkt ("theoretische Mitte" in den oben zitierten Tabellen) der individuellen Gütemaßstäbe nehmen wir hier - in Konsequenz der oben getroffenen Operationalisierungen - denjenigen Punkt auf dem jeweiligen Differenzen-Spektrum an, der (unter Berücksichtigung der Vorzeichen) die Mitte zwischen größter und kleinster Differenz eines Wertungsmaßstabes markiert, die arithmetische Mitte jedes individuellen Gütemaßstabes also.

Bei der Analyse der Daten fällt zu dem hier im Vordergrund der Betrachtung stehenden Aspekt als erstes auf, daß praktisch (bis auf eine Ausnahme) alle subjektiven Nullpunkte negative Differenzen beinhalten, d.h. in diesen gesamten Fällen war stets der mitgeteilte Sollwert des Anspruchsniveaus höher als der Punkt "beginnender Zufriedenheit".

Beispiel:

Der von der Vp als Anspruchsniveau mitgeteilte Zielwert lautet "40". Ein System-Mittelpunkt bei "-5" wird angenommen, wenn etwa im Zwei-Kategorien-Versuch bei dieser Vp die Zäsur zwischen "gut"- und "schlecht"-Urteilen bei einer Differenz von -5, also bei einem Punktwert von 35 zu lokalisieren ist. Im Drei-Kategorien-Versuch würde entsprechend über der Differenz "-5" das Urteilsmaximum der "mittel"-Kategorie liegen.
Umgekehrt würde ein Mittelpunkt von "+5" bedeuten, daß bei einer Vp z.B. im Zwei-Kategorien-Versuch auf dem Hintergrund eines Zielwertes von 40 Punkten die "Schnittstelle" zwischen "zufrieden" und "nicht zufrieden" bei einem Leistungswert von 45 Punkten liegt.

In der folgenden Tabelle sind die derart beschriebenen individuellen System-Mitten aus den 19 Versuchen der letzten Versuchsreihe zusammengestellt:

Vp	Versuchsdurchführung 1	2	3	4
LT	- 1,5	+ 1,0	- 3,0	- 3,5
RR	- 5,0	- 8,0	- 6,0	- 0,5
GM	- 8,0	- 4,5	- 6,5	- 4,0
ND	- 7,5	- 4,0	-10,5	-
JH	-13,5	- 9,5	-16,5	-12,0

Tabelle 4: Individuelle System-Mitten im Pfeilwurfversuch, ausgedrückt in Istwert-Sollwert-Differenzen.

Die einzige positive Mittel-Lage kommt bei Vp LT im zweiten Versuch vor, bei einer Versuchsperson, für die auch sonst kleine Istwert-Sollwert-Differenzen kennzeichnend sind. LT setzt seine Zielwerte (Anspruchsniveau) praktisch immer so, daß sie durch die Wurfleistungen fast ebenso oft überboten wie unterschritten werden; die Differenzwerte sind auf diese Art fast ebenso oft positiv wie negativ.
Ganz anders das Zielsetzungsverhalten etwa von JH: er setzt seine Ziele immer so, daß sie knapp über der augenblicklichen Maximal-Leistung liegen; die Wurfleistungen müssen also praktisch immer mehr oder weniger deutlich unter dem Ziel liegen, denn bereits bei einer einzigen dichten Annäherung an das Ziel erhöhte JH den Zielwert - im Gegensatz zu LT, der sein Ziel jeweils mehrere Male übertraf, bevor er es erhöhte.
Die übrigen Vpn nehmen hier zwischen LT und JH verschiedene Zwischenstellungen ein; sie arbeiten mit begrenztem Risiko, indem sie ihre Ziele so hoch setzen, daß sie sich voll einsetzen müssen, daß sie aber auch eine echte Erfolgschance

haben. LT geht demgegenüber in diesem Sinne fast gar kein Risiko ein; JH - in ganz anderer Art - auch nicht, da seine Zieldiskrepanz so groß ist, daß Erfolge - wenn man darunter ein Erreichen oder Übertreffen des Ziel(wert)es versteht - objektiv recht unwahrscheinlich sind. (JH tut hier imgrunde etwas ähnliches, wie wir es vor allem aus unserer ersten Versuchsreihe berichteten: er hält sich einen Weg offen, um objektive Mißerfolge in Erfolge umzudefinieren, denn - natürlich! - kann man nicht immer seine maximale Leistung bringen.)

Ohne hier zu sehr vereinfachen zu wollen, möchten wir doch darauf hinweisen, daß diese individuellen Nullpunkt-Lagen der Beurteilungssysteme für die eigene Wurfleistung uns sehr stark an die vor allem mit HECKHAUSENs Namen verknüpften Begriffe "Hoffnung auf Erfolg" und "Furcht vor Mißerfolg" denken lassen. Die Vpn RR und GM wären danach (vgl. u.a. HECKHAUSEN 1963b) wohl als eher "erfolgsmotiviert", LT hingegen als eher "mißerfolgsmotiviert" zu klassifizieren. In dieser Terminologie könnte man JH als "konfliktgeladen mißerfolgsmotiviert" bezeichnen - aber diese Verbindung soll hier nur angedeutet werden, wir wollen die Aussagekraft unserer Daten nicht überstrapazieren.

Festzuhalten bleibt hingegen, daß wir selbst bei unseren relativ wenigen Versuchspersonen interindividuell grundsätzlich verschiedene System-Nullpunkte bei der Beurteilung der eigenen Wurfleistung feststellen. Ohne Zahlenverhältnisse überzuinterpretieren, können wir wohl als Tatsache hinnehmen, daß Zufriedenheit mit der eigenen Leistung nicht erst beim Erreichen des Zielwertes beginnt, sondern bereits bei mehr oder weniger "schlechteren" Leistungen; System-Mitte im Sinne eines Bewertungs-Nullpunktes und Zielwert im Sinne des Anspruchsniveaus sind nicht identisch, sondern das Anspruchsniveau als Zielhöhe, als "Schwierigkeitsgrad eines Zieles,

wonach eine Person strebt" (LEWIN 1963), bleibt praktisch in jedem Falle Ziel - relativ gut erreichbar im allgemeinen, aber es erfordert Leistung und Einsatz. Der Nullpunkt der Leistungsbeurteilung hingegen liegt normalerweise in einem niedrigeren Bereich der Leistung, "reagiert" aber sehr fein, d.h. er verschiebt sich recht prompt in Abhängigkeit von Leistungsänderungen, wie wir vor allem am Beispiel der Vp HO im Reaktionszeitversuch gesehen haben.
Unterstützt wird diese Interpretation zusätzlich durch einen Befund, auf den wir bereits oben bei der Erörterung der Zwei-Kategorien-Versuche hinwiesen: Dort waren die beobachteten Zäsurwerte - in jenem Falle identisch mit den Systemmitten - bei den Urteilsverteilungen aller Versuchspersonen größer (weniger im negativen Bereich) gewesen als die zu erwartenden Werte. Wir hatten dazu gemeint, dies weise darauf hin, daß die Vpn mehr von sich erwarteten als sie dann leisteten, wodurch der Anteil der "schlecht"-Urteile (bei "ehrlicher" Bewertung) etwas größer wäre als der der "gut"-Wertungen. Analysiert man unter diesem Gesichtspunkt die Daten der übrigen Versuche, so zeigt sich kein entsprechend klares Bild. Ablesbar ist eine analoge Tendenz lediglich für die "negativ"-Seiten der individuellen Wertungsspektren - in den "positiven" Abschnitten sieht es genau umgekehrt aus. Es sprechen somit wohl auch die Ergebnisse aus dieser Datenbetrachtung für die kolossale Bedeutung des Anspruchsniveaus als Zielwert. Wenn wir dies hier aus diesen Daten nicht vollkommen exakt nachweisen können, so mag daraus ersichtlich sein, daß hier noch andere Momente eingehen, z.B. die Umfänge der Mittelbereiche. Eine Überprüfung der einzelnen Systemumfänge, von denen - vor allem durch die hier notwendige Operationalisierung - die Bewertungsnullpunkte mit abhängig waren, ergibt insgesamt über die Versuche hinweg gute interindividuelle Übereinstimmung. Am ehesten auffallend sind dabei die beiden folgenden

Befunde: Zum einen urteilt die Vp GM praktisch in allen Versuchsdurchführungen aufgrund des größten Wertungssystems; andererseits kommen bei der Mehrzahl der Vpn im jeweils ersten Versuch die umfangreichsten Systeme vor. Beide Ergebnisse erklären wir mit dem unterschiedlichen Leistungsumfang, der - wenn unsere Hypothesen richtig sind - unterschiedliche Wertungsmaßstäbe bedingen muß. Insofern wäre dies eine Bestätigung für die Aussagekraft unserer Ergebnisse, speziell auch zur Lage der System-Mitten.

5.3.5. Anspruchsniveau-Setzungen und Anspruchsniveau-Verschiebungen

Wir haben bereits im letzten Absatz implizit notwendigerweise die individuellen Eigenheiten von Versuchspersonen bei Anspruchsniveau-Setzungen im Pfeilwurf-Versuch mitbehandelt, dargestellt am Beispiel der Vpn unserer letzten Versuchsreihe. Da hier zudem keine "Motivationsanalyse der Anspruchsniveau-Setzung" (HECKHAUSEN 1955) beabsichtigt ist, wollen wir auf diesen Bereich hier nur noch soweit eingehen, wie es für unsere Fragestellung relevant scheint.
Es wurde bereits erörtert, daß die Anspruchsniveaus (ANs) unserer Vpn - hier operational definiert als Differenz zwischen mitgeteiltem Zielwert und Nullpunkt der Leistungsbeurteilung - verschieden hoch sind, daß sich dabei möglicherweise auch Hinweise auf die Erfolgs- resp. Mißerfolgsmotivierung sensu HECKHAUSEN für die einzelnen Vpn ergeben. Das ganze Augenmerk soll daher hier nicht den anfänglichen AN-Setzungen, sondern den aus den Leistungen resultierenden Verschiebungen der verschiedenen ANs gelten, behandelt wiederum primär an den quantitativ besonders aussagekräftigen Daten der letzten Versuchsreihe.
Wir stellen dabei als erstes fest, daß in allen 19 Versuchen dieser Serie insgesamt nur 83 AN-Verschiebungen vorkommen - wesentlich weniger auf jeden Fall, als unsere Vpn hätten vor-

nehmen müssen, wenn man nach jedem "Erfolg" oder "Mißerfolg" eine Reaktion erwartet hätte - denn immerhin umfaßte jeder Versuch ja 50 Leistungsergebnisse.
Hinzu kommt, daß es sich bei den 83 AN-Verschiebungen um 68 Erhöhungen und 15 Verminderungen handelt; von den letzteren entfallen allein 12 auf eine einzige Vp (ND), davon wiederum 9 auf einen einzigen Versuch (2.). Wir sehen also, daß bis auf diese Ausnahme praktisch nur Anspruchsniveau-Erhöhungen vorgekommen sind - was gleichzeitig auch die absolut niedrige Zahl der AN-Änderungen erklärt, denn Leistungssteigerungen erfolgen nicht so schnell und wir haben ja gesehen, daß im allgemeinen der Zielwert nur vergleichsweise knapp über dem Leistungsstand liegt.
Wir wollen angesichts dieser Ergebnisse unsere Interpretation hier auf zwei Punkte beschränken: auf das generelle Verhalten unserer Versuchspersonen bei AN-Änderungen und auf das Verhalten der Vp ND dabei im Detail.
Zum generellen Zielsetzungsverhalten würden wir zunächst einmal - im Sinne HECKHAUSENs - wesentlich häufigere Reaktionen unserer Vpn erwarten, im "typischen" Falle etwa jeweils eine Erhöhung des Zielwertes nach Erfolg und eine Verminderung nach Mißerfolg. Sollte diese These dennoch stimmen - und es spricht immerhin eine ganze Reihe von Ergebnissen aus der Anspruchsniveau-Forschung dafür - so würde das bedeuten, daß in unseren Versuchen nicht notwendigerweise jedes Erreichen oder Übertreffen des Zielwertes auch als Erfolg, nicht jedes Nichterreichen als Mißerfolg erlebt wird. Wir haben oben bereits gesehen, daß mindestens der Mißerfolg nach "unten" hin eine Grenze hat - und die AN-Forschung (seit HOPPE 1930) lehrt, daß dies auch für den Erfolg nach "oben" hin gilt, wofür wir allerdings keine Hinweise finden konnten. Zusätzlich stellen wir nun noch hier fest, daß auch innerhalb des entsprechenden Bezugssystems (im

Sinne des Bewertungsrahmens) wenigstens der Mißerfolg erst dann überhaupt als solcher erlebt wird, wenn er ein bestimmtes Ausmaß und/oder eine bestimmte Häufigkeit überschritten hat. Ein besonders anschauliches Beispiel hierfür bietet das Änderungsverhalten der Vp ND in ihrem zweiten Versuch, das in der folgenden Tabelle als Verlaufsreihe wiedergegeben ist:

Durchgang	E = Erfolg / M = Mißerfolg									
1-10	E!	M	E!	M!	M	M!	E!	M	M	M!
11-20	M	M!	M	E!	M	M	M	E!	M	M!
21-30	M	M	M	M!	E!	M!	M	M	M	M
31-40	M	E!	M	E	M	M!	M	M!	M	E!
41-50	M	M!	M	M!	M	E!	E!	M	M	M

! = Änderung des Anspruchsniveaus
_ = Zieldiskrepanz von mindestens 15 Punkten

Tabelle 5: Änderungen des Anspruchsniveaus im Pfeilwurf-Versuch (Vp ND, 2.Versuch)

Wir stellen fest, daß bis auf eine Ausnahme (im Durchgang 34) ein einziger Erfolg, ein einziges Erreichen oder Übertreffen des Zielwertes genügt, um das Anspruchsniveau zu erhöhen - zu einer Verminderung, zu einem Mißerfolgserleben im Sinne von HECKHAUSEN hingegen sind große (über 15 Punkte) bzw. mehrere Abweichungen erforderlich.
Dies deckt sich mit den entsprechenden Ergebnissen aus unseren früheren Versuchen und würde die oben schon ausgesprochene Vermutung bestätigen, daß ein Mißerfolg im Sinne des Nichterreichens des Zieles zunächst einmal nicht als "Mißerfolg" erlebt wird, sondern als "Ausbleiben des Erfolges" - ähnlich wohl wie beim Glücksspieler. Es müssen erst quantitative und/oder Wiederholungsaspekte hinzukommen, um den

Mißerfolg zu einem solchen werden zu lassen, um zu einer Verminderung des AN Anlaß zu geben. Wenigstens der Mißerfolg - für den Erfolg konnten wir dergleichen nicht feststellen - hat also offenkundig nicht nur eine "Maximalgrenze", jenseits derer er nicht mehr als solcher erlebt wird, wie wir bereits zu Beginn unserer Untersuchung gesehen haben und wie es aufgrund der Ergebnisse der AN-Forschung schon selbstverständlich ist, sondern er hat auch deutlich eine "Minimalgrenze", jenseits derer er noch nicht als solcher beurteilt wird, er wird also kognitiv ganz augenscheinlich "in Grenzen" gehalten.
Gleichzeitig ist diese verschiedenartige Einschätzung und Behandlung von Erfolg und Mißerfolg durch unsere Vpn aber auch eine Bestätigung dafür, daß der Nullpunkt des Bezugssystems für die eigene Leistung notwendigerweise mehr oder weniger unter dem Zielwert, unter dem Anspruchsniveau liegt. Würden unsere Versuchspersonen nämlich Erfolg und Mißerfolg - im definierten Sinne als Erreichen oder nicht-Erreichen des Zieles - in gleicher Weise erleben und entsprechend reagieren, so müßten beide Werte zusammenfallen.

5.4. Zusammenfassung

Die in diesem Kapitel geschilderte Versuchsreihe wurde im wesentlichen aus zwei Gründen durchgeführt: Zum einen sollten ihre Ergebnisse die in den früheren Versuchen gewonnenen Befunde, die sich vor allem durch ihre qualitative Vielfalt ausgezeichnet hatten, aus einer stringenteren Versuchsanordnung heraus bestätigen. Wir wollten sehen, ob sich die gleichen Verhältnisse, die sich oben bereits angedeutet hatten, auch unter neutraleren und anregungsärmeren Leistungsbedingungen nachweisen ließen, und veränderten Versuchsanordnung und Versuchsdurchführung entsprechend. Auf der anderen Seite wollten wir in dieser Serie durch eine relative Standardisie-

rung der Versuche (eben aufgrund der eingeschränkten Bedingungen) in dieser weitgehend explorativ angelegten Studie einen ersten Schritt in Richtung Quantifizierung gehen. Ziel war, einen möglichen Weg zur näheren Erforschung der Bezugssystemstrukturen auf dem Gebiete des Leistungsverhaltens resp. der Leistungsbeurteilung aufzuzeigen.

Wir haben gesehen, daß zunächst einmal auch in dieser Versuchsanordnung noch die Versuchspersonen "im Bedarfsfalle" ohne Schwierigkeiten genügend Möglichkeiten ersinnen können, um ihre Mißerfolgsquoten zu verkleinern, indem sie ausweichen auf andere Beurteilungsmaßstäbe und damit in andere Bezugssysteme. Es scheint dies nicht nur ein "Täuschungsmanöver" gegenüber dem Versuchsleiter und damit experimentell bedingt zu sein - wenngleich wir die Bedeutung sozialpsychologischer Komponenten gerade auf dem Gebiet der Leistung und der Leistungsbewertung nicht unterschätzen -, sondern wirklich auch eine Art bereitwilligen Selbstbetruges, wie wir aus den Interviews und auch aus Selbstversuchen entnehmen. Vermutlich gibt es gerade in diesem Person-Umwelt-Bezug des Leistungsverhaltens derart starke psychohygienische Tendenzen, daß sie die bereits nachgewiesenen Systemgrenzen nach "unten" hin, zur subjektiv "schlechteren" Leistung hin also, verstärken, sie nur noch absoluter machen.

Bei der Aufarbeitung der Daten dieser Versuchsreihe nahmen wir aufgrund der bisherigen Erfahrungen einige Operationalisierungen vor, vor allem hinsichtlich der System- und Kategorialgrenzen. Diese Notwendigkeit ergab sich hier aus der Art der Versuchsdurchführung, anders als dies bei den früheren Versuchen der Fall gewesen war.
Wir stellten fest, daß vor allem interindividuelle Unterschiede hinsichtlich der Systemumfänge und der Systemmitten bezüglich der Beurteilung der eigenen Wurfleistung in den

Versuchsergebnissen deutlich werden, die auf der einen Seite auf verschiedenen Leistungsumfängen beruhen, auf der anderen Seite aber auch verschiedene Zielsetzungs-Verhaltensweisen erkennen lassen. Intraindividuelle Unterschiede sind demgegenüber zweitrangig, sie lassen sich im wesentlichen - und dafür spricht unsere Datenanalyse - mit Leistungsstabilisierungen und "Einkennungen" in die Versuchsanordnungen erklären.

Es wurden Versuche zur Bewertung der eigenen Leistung im Pfeilwerfen nach 2, 3, 4 und 5 Kategorien durchgeführt. Wir konnten dabei - wenn auch nicht in allen Fällen - Bezugssystemstrukturen erkennen, die belegen, daß eine Bewertung der eigenen Wurfleistung nach verschiedenen Kategorien-Anzahlen in der Weise möglich ist, daß dem Beurteilungsprozeß offenkundig kognitive Strukturen zugrunde liegen, die entsprechend ausdifferenziert sind. Es wurde wenigstens für einige Versuchspersonen klar - und damit erst eigentlich ein sicherer Existenznachweis geliefert - daß für die Beurteilung der eigenen Leistung (etwa im Pfeilwerfen) die gleichen Gesetzmäßigkeiten gelten, wie sie nach WITTEs Theoremen zu erwarten waren - wobei diese aufgrund von Gegebenheiten gewonnen wurden, die sich über physikalische Metriken abbilden lassen. Diese Übereinstimmung gilt für die Bereichsumfänge und für die Lage der Zäsuren zwischen den Kategorial-Bereichen.

In einigen Fällen lassen die Urteilsverteilungen - bezogen auf die jeweiligen Istwert-Sollwert-Differenzen - darauf schließen, daß unsere Vpn kognitiv weniger Kategorialbereiche durchstrukturiert haben als sie verschiedene Begriffsstufen verwenden; sie können verbal z.B. durchaus vier Begriffsklassen zur Wertung verwenden und dennoch nur drei Bereiche wirklich ausdifferenziert haben - etwa, indem die beiden mittleren Wertungsklassen eigentlich zusammenfallen.

Dies mag an der teilweise ungenügenden Vertrautheit mit speziell dieser Art der Leistungsmessung und der daraus resultierenden Leistungsauseinandersetzung gelegen haben - wir verstehen es jedoch letztlich als Bestätigung der zuerst geschilderten Befunde: Die kognitive Struktur, die der Beurteilung der eigenen Leistung zugrunde liegt, ist ganz offenbar individuell derart klar ausgeprägt, daß sie selbst trotz begrifflicher Unsicherheiten und Unschärfen deutlich wird. Zusätzliche Hinweise auf die Bezugssystem-Genese bei der Beurteilung der eigenen Leistung erhielten wir in dieser Versuchsreihe vor allem durch den Fünf-Kategorien-Versuch der Vp LT.

6. Schlußbetrachtung

6.1. Zusammenfassende Erörterung

Auf der Grundlage der bisherigen Untersuchung von mnestisch stabilisierten Bezugssystemen formulierte WITTE (1960a) im Rahmen seiner Bezugssystemtheorie, die auf das Prinzip der Strukturierung nach Ähnlichkeit gegründet ist, eine Reihe allgemeingültiger Gesetzmäßigkeiten, von denen wir die für das daraus resultierende Modell hervorragendsten in der Formulierung von FORNER (1974) bereits zu Beginn dieser Arbeit zitiert haben.

Dieses Gesamt von Hypothesen wurde auch von FORNER im Rahmen der vorliegenden Befunde mathematisch-statistisch überprüft. Die Verfasserin gelangte dabei zu folgenden Ergebnissen:

a. zu der von WITTE postulierten Normalverteilung von relativen Bereichsumfängen, absolutem Systemumfang und Systemenden:

   (1) "Die Verteilung der Systemenden und des Systemumfanges werden durch das Nicht-Verwenden echter Systembegrenzungen ... empfindlich beeinträchtigt." (Trendanalyse)

   (2) "Die Verteilung der relativen Bereichsumfänge wird nicht von dem Vorhandensein echter Systembegrenzungen beeinträchtigt, gleichgültig ob keine, nur eine oder beide Begrenzungen vorhanden sind." (Trendanal.)

b. zur Unabhängigkeit (Nullkorrelation)

   - von relativem Bereichsumfang und Systemumfang;
   - von Systemanfang und Systemende sowie
   - von Systemenden und Bereichsumfang:

   (3) "Die Mittelwerte der Korrelationskoeffizienten zeigen deutlich eine Tendenz zur Nullkorrelation, also zu Unabhängigkeit."

   (4) "Nur relativ wenige Koeffizienten zeigen gesicherten Zusammenhang an, die meisten Koeffizienten lagen innerhalb der Streugrenzen $\pm 1\ \sigma$ (Mittelwert=0)."

(5) "Es trat ein Konstanteneffekt durch die nicht verwendeten Systemenden auf ..."

c. zum Streuungsvergleich:

(6) "Beide Methoden führen zu dem gleichen Ergebnis, nämlich daß die intraindividuelle Streuung signifikant kleiner ist als die interindividuelle Streuung."

All diese Ergebnisse beziehen sich auf Bezugssysteme von Mannigfaltigkeiten, die über physikalische Metriken (cm, g, Hz usw.) eingeordnet werden können.
Bereits 1966 hatte KELLNER eine Untersuchung angestellt an einer bezugssystem-bildenden Mannigfaltigkeit, die nicht derart auf physikalische Meßgrundlagen bezogen werden kann, nämlich an sportlerischen Leistungen, dargestellt an der gegenseitigen Beurteilung im Handballspiel von Polizeischülern. Die Ergebnisse seiner Studie "legen bei Annahme (1) normalverteilter Leistungen und - wie im vorliegenden Fall vermutlich gegeben - (2a) realistischer Wahrnehmung derselben sowie (2b) objektiver Beurteilung eine Deutung gemäß der gleichgroße phänomenale Kategorialabschnitte unterstellenden Theorie mnestisch stabilisierter Bezugssysteme nahe", wie WITTE (1971) formuliert.
SIEVERS (1970) wollte einen Schritt weitergehen und - wie von HECKHAUSEN (1963a) bereits früher angeregt - die Gedanken der Bezugssystemforschung gezielt auf die Untersuchung von Gütemaßstäben anwenden, auf motivationspsychologische Fragen also. Dabei geht er einerseits jedoch ganz auf formale Aspekte aus und muß so inhaltliche Fragen vernachlässigen, andererseits verwendet er im Grunde die KELLNERsche Versuchsanordnung: auch er läßt Schüler einander pauschal beurteilen und zwar nach "Güte" im Fußballspiel und nach Lernfähigkeit.

Wir wollten demgegenüber hier in dieser Studie - ganz auf dem Hintergrund der HECKHAUSENschen Gedanken - auf die Bewertung der <u>eigenen</u> Leistung abheben, denn hier dürfte eine

Untersuchung von Bezugssystemstrukturen auf dem Gebiete der Leistungsmotivation als einer Auseinandersetzung mit Gütemaßstäben zuvörderst am Platze sein.
Dabei hielten wir es für notwendig, phänomenologisch-explorativ vorzugehen, denn auf der einen Seite haben wir es hier zwar mit einem Gebiet zu tun, dessen Strukturen und Dynamiken uns allen wohl vertraut sind - "wissen" wir doch bei all unseren Leistungen sehr genau, ob wir mit den Ergebnissen "zufrieden" oder "nicht zufrieden" sind, ob wir sie (und damit auch meist uns selbst) "gut" oder "schlecht" finden, sofern sie nur für uns Relevanz haben.
Andererseits aber haben wir uns dabei hier auf ein Untersuchungsfeld begeben, das - wenigstens vorläufig - in der Bezugssystemforschung einzigartig ist: Nicht nur, daß die Mannigfaltigkeit "Leistung", von der wir annehmen, daß sie bezugssystembildend sei, in dem von uns gemeinten Sinne nicht über physikalische Metriken abtragbar ist - sie ist noch nicht einmal für alle Versuchspersonen identisch, d.h. jede Vp beurteilt eine andere Mannigfaltigkeit, nämlich ihre eigene Leistung.
In allen bisherigen Versuchen war das "Versuchsmaterial" konstant, sei es eine Serie von Bleistiften oder Gewichten, ein Pool von Klassenkameraden (KELLNER, SIEVERS) oder die Systematik von Lottogewinnen (SKROBLIN) - wir mußten demgegenüber mit rein individuellen "Mannigfaltigkeiten" arbeiten, denn nur die Bewertung der <u>eigenen</u> Leistung - auf welcher Kontrastfolie auch immer - kann letztlich Relevanz für unser Handeln haben und damit verhaltenssteuernd sein.

Wir haben dabei in unserer experimentellen Studie über die Bewertung der eigenen Leistung - dargestellt primär am Beispiel des Pfeilwurf-Spiels, daneben auch an Reaktionszeitversuchen - herausgefunden, daß auch für die Beurteilung der Istwert-Sollwert-Diskrepanzen im Person-Umwelt-Bezug

"Leistung" sensu HECKHAUSEN die WITTEschen Bezugssystem-Annahmen zu gelten scheinen.
Bereits ein erster Blick etwa auf die graphischen Darstellungen unserer Ergebnisse zeigt recht hohe Übereinstimmung mit den Häufigkeitsverteilungen, wie wir sie aus früheren Bezugssystem-Versuchen gewohnt sind. Die Lage der Zäsuren zwischen den Kategorialabschnitten zeichnet sich ebenfalls durch teilweise nur höchst geringfügige Abweichungen von den erwarteten Stellen aus.
Sicherlich gilt dies nicht für all unsere Versuche, aber hier sollte ja zunächst einmal überhaupt ein Existenznachweis geführt werden. Auf diesem Hintergrund dürfte die Anzahl der hypothesengerecht ausgefallenen Versuchsergebnisse auf jeden Fall überzufällig hoch sein - wieweit bei den übrigen etwaige Fehler unserer Versuchsanordnung maßgeblich waren, wird noch zu diskutieren sein.

Wenn wir davon ausgehen - und wir können bezüglich unserer Versuchsanordnungen eigentlich keine stichhaltigen Gegenargumente finden -, daß Pfeilwurf-Leistungen und Reaktionszeiten normalverteilt sind; wenn wir weiterhin unterstellen, daß unsere Versuchspersonen ihre Leistungen leidlich objektiv (oder doch wenigstens konstant) beurteilt haben, dann liegt aufgrund unserer Befunde der Schluß nahe, daß WITTEs Theorem von den im Erlebnis der Vpn gleichgroßen Kategorialabschnitten auch für die Bewertung eigener Leistungen Gültigkeit hat. Wenn wir auch interindividuell aufgrund der großen Streubreite der Leistungs- und Wertungsbereiche kaum Aussagen machen können, so sind doch die intraindividuellen Ergebnisse insgesamt recht befriedigend.
In aller Deutlichkeit gilt dies für den größten Teil der Zwei- und der Drei-Kategorien-Beurteilungen, aus den Versuchen mit vier und mit fünf Bewertungsstufen lassen sich lediglich Tendenzen ablesen. Wir können unsere Daten - aufgrund

Eigenart des "Materials" - nicht über mehrere Versuche oder Versuchspersonen hin zusammenfassen. Dennoch zeigen sich in der Mehrzahl der Fälle klare Bezugssystemstrukturen in der Form, wie wir sie in unseren Hypothesen erwartet hatten.

Es scheint dabei aber im allgemeinen bei der Auseinandersetzung mit der eigenen Leistung nicht nur einen Gütemaßstab zu geben, nach dem wir unsere Leistungsergebnisse beurteilen, nicht nur ein Bezugssystem für Erfolg oder Mißerfolg. Vielmehr tendieren wir dazu - sofern wir die Freiheit bzw. den Spielraum dazu haben - im Falle des Mißererfolges in andere Systeme überzuspringen, andere Maßstäbe anzulegen, die uns das Eingeständnis des Mißerfolges ersparen oder wenigstens erleichtern sollen. Im Versuch kann dieser Neigung allenfalls dadurch teilweise begegnet werden, daß die anzulegenden Maßstäbe möglichst präzise definiert werden.
Für den Fall des Erfolgserlebens konnten wir dergleichen nicht feststellen - denkbar wäre es allerdings, etwa im Setzen grundsätzlich neuer, höherer Ziele nach dem Erreichen der bisherigen.

Die die jeweiligen Gütemaßstäbe determinierenden Soll-Lagen können sich als Normal- oder Ideallagen sensu HECKHAUSEN (1963a) darstellen. Dabei sind nach unseren Ergebnissen zunächst ganz offenkundig "sachimmanente Prägnanzstrukturen" der zu bewältigenden Aufgabe, dann "Erfahrungsniederschläge" und schließlich (in eben dieser Reihenfolge) "Norm-Vorstellungen" im Sinne sozialer Bezüge maßgeblich.
Zunächst fallen tatsächlich die Aufgaben-Kategorien leicht/schwierig mit den Urteilsbegriffen gut/schlecht bei der Ausprägung von Maßstäben zusammen, bevor dann die eigene Erfahrung in den Vordergrund tritt; d.h. in einem Falle, wo erst eine Einkennung in eine ganz spezielle Leistungsart erfolgen muß, werden sozusagen zunächst von der Aufgabe her theoretische Erfahrungswerte festgelegt, die dann durch die eigene

Leistung modifiziert werden. Soziale Bezüge schließlich bringen bei der Leistungsbeurteilung relative Wertungen mit sich und damit letztlich die Abkehr von den hier im Vordergrund stehenden Absolutbeurteilungen. Die eigentlichen Gütemaßstäbe zur Bemessung der eigenen Leistung können so - falls aus psychodynamischen Gründen notwendig - vergleichsweise (im wahrsten Sinne des Wortes!) gut verlassen werden, es gelten ganz andere Gesetze, wie wir bereits in unserer ersten Versuchsreihe sehr anschaulich feststellten.

Vor allem in den beiden letzten Versuchsreihen (Reaktionszeiten und Dart-Spiel) wollten wir zur besseren Untersuchung der Bezugssystemstrukturen ganz auf die Erfahrungskomponente abheben, indem wir auf der einen Seite den Einfluß sachimmanenter Prägnanzstrukturen einzuschränken versuchten und auf der anderen auch sozial relevante Aspekte nach Möglichkeit ausschalteten.
Wir geben uns hier wohl keiner Täuschung hin, wenn wir der Anwesenheit des Versuchsleiters gerade auf dem Gebiet der Leistung und der Leistungsbeurteilung Gewicht zumessen - HECKHAUSEN etwa weist 1955 auf diesen Punkt hin, auch BOCHOW hebt (in Anlehnung an die HOPPE-Versuche) die Bedeutung dieses Aspekts hervor. Wir glauben dennoch, im Sinne unserer Versuchsintention hier nicht ganz erfolglos geblieben zu sein. Gerade die Wahl von Versuchspersonen aus dem persönlichen Bekanntenkreis - die wir zu Beginn dieser Arbeit begründeten - in Verbindung mit einer betont neutralen Versuchsatmosphäre dürfte hier den angestrebten Zweck erreicht haben. Außerdem: keine Versuchsperson der letzten beiden Versuchsreihen wußte, wie gut (oder schlecht) der Versuchsleiter in den geforderten Leistungen war.
Wenn also primär Erfahrungsniederschläge maßgebend sind für die Beurteilung der eigenen Leistung, dann haben wir tatsächlich Normallagen "nach Art sich wandelnder Adaptations-

niveaus" (HECKHAUSEN) gefunden. Etwa bei den Reaktionszeitversuchen genügten bereits relativ wenige Durchgänge, um in Abhängigkeit von Leistungsänderungen Verschiebungen des Nullpunkts des relevanten Gütemaßstabes nachzuweisen. Dennoch dürfte der jeweils realisierte Leistungsstand hier oft nicht allein maßgebend gewesen sein, Gespräche einzelner Vpn über ihr "Abschneiden"waren sicher nicht ganz zu verhindern.
In jedem Falle aber können wir wohl nach unseren Versuchen davon ausgehen, daß der Existenznachweis geführt scheint, daß die WITTEschen Bezugssystem-Theoreme auch für die Bewertung der eigenen Leistung gelten, und zwar sowohl hinsichtlich der Normalverteilung als auch der Unabhängigkeit als auch der Streuung. Dies gilt nicht nur für die direkten Leistungsbeurteilungen unserer Vpn, sondern ganz deutlich auch für die von ihnen implizit vorgenommenen Bewertungen der Istwert-Sollwert-Diskrepanzen, auf die es uns ja eigentlich ankam - d.h. in der Struktur der Leistungsbeurteilungen läßt sich ganz klar der Bezug zur Zielsetzung ablesen, unsere Ergebnisse gelten also tatsächlich für die Ebene der Zufriedenheit bezüglich der eigenen Leistung.

Einige Modifizierungen scheinen dennoch angebracht:

Zunächst einmal hatten wir - vor allem aufgrund unserer Beobachtungen zum Vorgehen der Versuchspersonen - Grund zu der Annahme, daß die untersuchten Sollwert-Istwert-Diskrepanzen als Differenzen aufzufassen seien, daß der WEBER-Bruch demzufolge zum Verständnis des Leistungserlebens und der Leistungsbeurteilung nicht herangezogen zu werden braucht. Unsere Ergebnisse haben diesen Eindruck insgesamt bestätigt - dennoch bleibt die Frage, ob dies nicht überwiegend an der Versuchsanordnung (Dart-Scheibe) gelegen hat, wo der Unterschied von 20 zu 25 Punkten phänomenal ebenso groß ist wie der von 40 zu 45 Punkten. Möglicherweise liegen die Verhältnisse bei anderen Anordnungen anders, z.B. bei Reaktionszei-

ten, wo tatsächlich der Unterschied von 14 zu 15 cs phänomenal ein anderer ist als der von 24 zu 25 cs. Wir konnten dies aufgrund der dort in unseren Versuchen zu geringen Umfänge der Leistungsspektren nicht befriedigend klären, entsprechende Versuchsanordnungen sind jedoch denkbar.

Wesentlich deutlicher ist aber geworden, daß Erfolgserleben und Mißerfolgserleben nicht - oder nicht nur - einfach zwei Enden eines Kontinuums sind, wenigstens nicht aus dem Blickwinkel der Bezugssystemforschung.
Wir haben festgestellt, daß die individuellen Systeme für die Beurteilung der eigenen Leistung klare Grenzen haben - wenigstens nach "unten" hin: es gibt "Mißerfolge", die so gewaltig sind, daß sie schon nicht mehr als solche erlebt werden. Zwar lehrt uns die Anspruchsniveau-Forschung, daß es analog auch "zu" große Erfolge gibt - aus unseren Ergebnissen ist dergleichen nicht feststellbar, möglicherweise aufgrund der speziellen Versuchsanordnungen:

(a) Reaktionszeiten von 3 oder 4 Centi-Sekunden, die ganz offenkundig keine "echten" Werte darstellen konnten, kamen nicht vor und ließen sich an dem verwendeten Gerät auch nicht manipulieren;
(b) beim Pfeilwurf kamen zwar "extrem" gute Teilergebnisse vor, sie konnten aber in jedem Falle als mindestens theoretisch im Bereich des möglichen liegend integriert werden - hier konkurrierte die Prägnanzstruktur der Aufgabe durchaus erfolgreich mit der Erfahrungsbildung.

Auf der anderen Seite muß ein Mißerfolg (und möglicherweise auch der Erfolg) nach unseren Versuchen aber auch eine bestimmte Mindestgröße haben, um als solcher gewertet zu werden: nicht jede positive oder negative Zieldiskrepanz ist notwendigerweise überschwellig.

Wir haben oben bereits HECKHAUSEN (1965) zitiert, der meint, ein Gütemaßstab zerfalle auf jeden Fall "in einen Teil, der

- wie ausgeprägt auch immer - Erfolg, und in einen anderen, der Mißerfolg bedeutet." Nun, unsere Ergebnisse unterstreichen dies, aber **unsere** Befunde wenigstens zum Mißerfolgserleben und zum auf Mißerfolg folgenden Verhalten legen doch die Frage nahe, ob dies alles sei. Warum sollte es nicht gleichzeitig auch Gütemaßstäbe, Bezugssysteme also, geben, die nur kleineren oder größeren Mißerfolg, nur kleineren oder größeren Erfolg beurteilen lassen? Oder - im Bilde - warum nicht sowohl

Erfolg ——— O ——— Mißerfolg

als auch, daneben

minimaler Mißerfolg —— maximaler Mißerfolg

und minimaler Erfolg —— maximaler Erfolg ?

In Begriffen der Zufriedenheit würde dies auf die Frage hinauslaufen, ob "nicht zufrieden" semantisch identisch ist mit "unzufrieden" oder ob es etwa die begriffliche Mitte darstellt zwischen "zufrieden" und "unzufrieden".
Wir halten hier derartige Partialsysteme nach unseren Ergebnissen für immerhin möglich, vielleicht ist der Gütemaßstab für Erfolg oder Mißerfolg wirklich nicht so "einfach" im Sinne von eindimensional.

Darüber hinaus spricht auf dem Gebiete der Beurteilung der eigenen Leistung einiges dafür - auch aus der Alltagserfahrung -, daß wir mit ein klein wenig korrigierten Gütemaßstäben arbeiten: auf der einen Seite gehen wir mit Erfolgen, mit "guten" Leistungen oft recht großzügig um, d.h. wir rechnen sie uns auch dann zugute, wenn sie eigentlich schon außerhalb des Systems liegen; auf der anderen Seite werden Mißerfolge recht bereitwillig außerhalb des Maßstabes gestellt, neutralisiert - wahrscheinlich schon allein aus psychohygienischen Gründen, nicht nur also - wenn überhaupt - im Sinne einer mehr oder weniger bewußten kognitiven Verschleierung.

Wir haben ganz augenscheinlich auch für die Bewertung unserer Leistungen präzise Gütemaßstäbe, handhaben sie aber zum Teil anders als unsere Systeme etwa für die Beurteilung von Bleistift-Längen oder Gewichten - und wenn sich dies allein darin äußert, daß wir auf Mißerfolge in unseren Zielsetzungen schwerfälliger reagieren als auf Erfolge, was auch für eine qualitative Eingrenzung unserer Mißerfolgsquoten gilt: es kommt vor, wie wir gesehen haben, daß Mißerfolge zunächst einmal als "Ausbleiben des Erfolges" erlebt werden, das Nicht-Erreichen einer angestrebten Leistung wird so gleichsam zum "Noch-nicht".

Insgesamt aber können wir - auf dem Hintergrund des Gesagten - in Beantwortung unserer Versuchsfragen aufgrund dieser experimentellen Studie festhalten:

(1) Auf dem Gebiete der Beurteilung der eigenen Leistung finden wir Gütemaßstäbe vor, die den Bezugssystem-Theoremen von WITTE mindestens nicht widersprechen. Vielmehr stellen wir Urteilsverteilungen fest, die hinsichtlich ihrer kategorialen Struktur so ausgebildet sind, wie wir es aus früheren Versuchen mit ganz anderen Mannigfaltigkeiten kennen.

(2) Es sind aus unseren Ergebnissen Bezugssystem-Änderungen in Form von Nullpunkt-Verschiebungen nachweisbar, die auf eine sehr schnelle Anpassung der jeweiligen Adaptationsniveaus schließen lassen.

(3) Wenngleich sich ein Großteil des Urteilsgeschehens auf der Basis von Normallagen sensu HECKHAUSEN vollzieht, also eigenen Erfahrungen jeweils adaptiert wird, so genügt dies dennoch nicht zum Verständnis. Vielmehr sind in jedem Falle auch - vor allem im Stadium der Einkennung in eine ganz bestimmte Aufgabe - sachimmanente Strukturen beteiligt, daneben auch Normvorstellungen sozialer

Genese. Bei den letzteren ist allerdings die Grenze der Absolutbeurteilung erreicht, vergleichende Wertungen treten in den Vordergrund.

(4) Sofern die Sachstrukturen für die Ausprägung der relevanten Soll-Lagen bestimmend sind, fallen die Aufgaben-Beschreibungs-Kategorien leicht/schwierig mit den Wertungskategorien gut/schlecht zusammen. Dieser Bezug wird jedoch mit zunehmender Erfahrung gelockert.

(5) Welcher Art bzw. Genese die verschiedenen Gütemaßstäbe auch sind - ihre Strukturen scheinen ähnlich, womit ein zusätzlicher Beleg für die Gültigkeit unserer Ergebnisse für das Gebiet des Leistungsverhaltens gegeben ist.

(6) Die sich in der interindividuell verschiedenen Art der Anspruchsniveausetzung spiegelnden Erwartungssysteme lassen durch die Stellungnahme unserer Versuchspersonen zu den jeweiligen Sollagen-Istlagen-Diskrepanzen, die sie implizit ja beurteilten, auf unterschiedliche Ausrichtungen der individuellen Motivationslagen rückschließen.

6.2. Kritische Rückschau

Wir haben in der vorliegenden Studie Neuland im Rahmen der Bezugssystem-Forschung betreten - insofern nämlich, als wir Bezugssystem-Eigenschaften von Mannigfaltigkeiten untersucht haben, die einerseits in Pauschalurteile gefaßt und andererseits - darüber hinaus - für die einzelnen Beurteiler verschieden waren, indem sie ganz individuelle Leistungen betrafen.

Für die aus Gründen der besseren Kontrolle zwangsläufige Untersuchung im Labor-Experiment ergab sich die Notwendigkeit der Operationalisierung der "eigenen Leistung". Wir versuchten dieses Problem zu lösen, indem wir mit zwei Versuchsan-

ordnungen arbeiteten, bei denen sich die Leistungen in Werten (Punkt- bzw. Zeitwerte) ausdrückten, nämlich mit dem Dart-Spiel und mit dem Reaktionszeit-Meßgerät. Beide "Apparaturen" machten durch ihren geradezu zwingenden Aufforderungscharakter eine Leistungsmotivierung der Versuchspersonen durch eigens gesetzte Impulse überflüssig; es kam sogar vor, daß einige Versuchspersonen - besonders in der ersten Serie - die Versuche erheblich über den geplanten Umfang hinaus ausdehnten.

Wenn dies ganz im Sinne unserer Versuchsabsichten lag, so stellten sich im Verlaufe der Untersuchung doch einige grundsätzliche Probleme heraus:

Ziel war bei der Auswahl der Versuchsanordnungen - neben der offenen Leistungsthematik - zunächst einmal gewesen, mit "Mannigfaltigkeiten" zu arbeiten, denen mnestisch stabilisierte Bezugssysteme zugrunde lagen - denn auf diese stützt sich ja die WITTEsche Forschungsausrichtung, auf sie beziehen sich die zitierten Theoreme. Wir hatten in diesem Sinne vermutet, unsere Versuchspersonen verfügten über hinreichend präzise Gütemaßstäbe sowohl für ihre Zielmotorik als auch für ihre Reaktionszeiten. Zwar bestätigte sich dies insgesamt, es ergab sich aber als Beeinträchtigung, daß dennoch immer eine Einkennung sensu HELLPACH in die jeweils spezielle Art der Leistungsfeststellung notwendig war.

Bei den Dart-Versuchen etwa arbeiteten wir ja mit verschiedenen Wurf-Scheiben, der handelsüblichen sowie einer eigens angefertigten. Dabei zeigte sich nun, daß der Einfluß sachimmanenter Prägnanzstrukturen (wie es z.B. HECKHAUSEN nennt) sich nicht primär auf Beurteilungsaspekte wie Abstand des Pfeiles vom Mittelpunkt der Scheibe oder vom Rande bezieht. Wäre dies so, dann hätte die Detail-Gestaltung der Scheiben keine derartige Rolle gespielt. Tatsächlich aber standen Zahlenwerte im Vordergrund, die die Bedeutung der verschie-

denen Wurf-Ringe repräsentierten - und da bedeutete ein Wert von "50" bei der ersten Scheibe nun wirklich etwas anderes als bei der zweiten, waren doch auf der ersten z.B. mit sechs Würfen 600 Punkte möglich, auf der zweiten nur 66. Zwar warf bei den verschiedenen Untersuchungsserien keine Versuchsperson auf beide Scheiben, eine Einstellung auf die jeweils vorgegebene Struktur war dennoch notwendig - vermutlich auch aufgrund von spontanen Operationalisierungstendenzen unserer Vpn: niemand würde sagen (und niemand es verstehen!): "ich habe so ein Stück von der Mitte weg geworfen", sondern man hat z.B. "eine 5" geworfen, geschossen o.ä.. Für unsere Fragestellung wäre dergleichen dennoch unerheblich, würde sich nicht hier gerade ausdrücken, wie der Leistungsbereich einer Vp bei ihr selbst abgebildet ist. Ähnlich war es im übrigen bei den Reaktionszeitversuchen: Zwar weiß jeder normalerweise aus seinen vielfältigen Alltagserfahrungen, wie schnell er "schalten" kann - dazusitzen jedoch, auf einen Signalton zu warten und diesen möglichst schnell durch Knopfdruck zu beantworten, ist etwas anderes, erfordert eine Einkennung.

Insofern mußten wir unsere ursprünglichen Vorstellungen über die mnestische Stabilität unseres Untersuchungsthemas sicherlich im Verlaufe der Versuchsdurchführungen modifizieren, mußten der jeweiligen Einübung mehr Raum geben als eigentlich vorgesehen war. Auf dieses Konto geht vermutlich der recht eigenartige Ausgang mancher Einzelversuche, wo z.B. zwar vier Kategorien zur Leistungsbeurteilung verwendet wurden, eigentlich aber nur drei Leistungsbereiche dadurch repräsentiert sind. Wir sehen dies auf der einen Seite durchaus als Versuchsfehler, daß nämlich letztenendes inadäquate Kategorien vorgegeben wurden: die Versuchspersonen sollten etwas leisten, was sie (noch?) nicht leisten konnten. Auf der anderen Seite aber sehen wir diese Ergeb-

nisse - und gerade sie - als bestätigenden Beleg auch für die übrigen Versuche, deren Urteilsverteilungen so sind, wie wir es erwartet hatten: sie sprechen nämlich unserer Ansicht nach klar dafür, daß verschiedene Leistungsbereiche klar kognitiv durchstrukturiert sind - so präzise, daß sie sich selbst auf dem Hintergrund inadäquater Kategorien abbilden, z.B. indem zwei Kategorialabschnitte zusammenfallen. Dies scheint der Forderung nach erlebnismäßig gleichgroßen Kategorialbereichen zu widersprechen, läßt sich aber wohl dennoch kaum von der Hand weisen. Teilweise war kaum zu übersehen, daß - um bei dem bereits angedeuteten Beispiel zu bleiben - die Vpn zwar verbal vier Kategorien verwendeten, weil diese vorgegeben waren, daß aber eigentlich zwischen "eher gut" und "eher schlecht" für sie kognitiv kein Unterschied war, daß diese Begriffe also letztlich für den gleichen Leistungsbereich standen.
Wir können somit wohl davon ausgehen, daß die Bezugssysteme für die Beurteilung der eigenen Pfeilwurf-Leistung oder Reaktionszeit nicht in dem Ausmaße mnestisch stabilisiert sind wie wir es aus den Versuchen mit Mannigfaltigkeiten auf sensorischer Grundlage gewöhnt sind - daß aber die Einkennung in die jeweils vorgegebene Art der Leistungsfeststellung und deren Beurteilung nur sehr wenig Zeit beansprucht. Dies kann dennoch nicht auf eine Erklärung unserer Befunde im Rahmen der Adaptationsniveau-Theorie HELSONs hinauslaufen, die - wie wir gesehen und erörtert haben - zu wenig von der Vielfalt der bei der Leistungsbeurteilung einschlägigen Gütemaßstäbe abdecken kann.

Ein anderer Aspekt des Leistungsgeschehens, der sich während unserer Untersuchung als problematisch herausstellte, ist die Frage der Leistungsvielfalt. Wir fanden vor allem in den ersten beiden Versuchsreihen, daß die Leistungsbereiche teilweise so schmal sind, daß überhaupt nur für zwei Kategorial-

abschnitte "Platz" ist, um hier eine Parallele zwischen Darstellung und kognitiver Strukturiertheit zu ziehen. Extrem formuliert:wenn es nur zwei Leistungsklassen gibt, ist es irrelevant, mit drei Urteilskategorien zu arbeiten. Wir versuchten diese Schwierigkeit durch entsprechende experimentelle Variationen zu überwinden. Daß wir dabei nur teilweise erfolgreich waren, mag hier an der Grundsätzlichkeit der Problematik gelegen haben. Es zeigte sich nämlich, daß mit dieser Frage eine charakteristische Eigenschaft speziell dieses Untersuchungsfeldes angesprochen ist: Je besser wir einen Leistungsbereich kennen, je mehr er mnestisch stabilisiert ist, desto konstanter sind (durch die größere Erfahrung) unsere Leistungen, desto kleiner ist das Leistungsspektrum, auf das sich die Urteile beziehen. Umgekehrt ließe damit ein breites Leistungs- und Urteilsspektrum, das sich in entsprechend viele Kategorialbereiche ausdifferenzieren könnte, auf mangelnde Erfahrung und somit relativ geringe mnestische Stabilisierung der entsprechenden Gütemaßstäbe schließen. Unsere Untersuchungen erbrachten insgesamt befriedigende Ergebnisse für den Fall der Beurteilung der eigenen Leistung nach zwei und nach drei Kategorien. Es scheint uns durchaus nicht sicher, daß es nur an ungenügender Erfahrung unserer Vpn mit den jeweiligen Versuchsanordnungen gelegen haben sollte, wenn für die Fälle von vier und von fünf Kategorien lediglich strukturelle Tendenzen nachweisbar waren. Vielmehr haben wir in unsere Überlegungen hier durchaus die Möglichkeit einbezogen, daß in diesem Punkte grundsätzliche Grenzen erreicht sind, die nach dem Gesagten in der Natur der Sache begründet lägen. Vielleicht gibt es bei der Einschätzung der eigenen Leistung ebenso wie bei ihrer Beurteilung wirklich nicht viel mehr Differenzierungsmöglichkeiten und -notwendigkeiten als "gut" und "schlecht" oder "Ziel erreicht" und "Ziel nicht erreicht" nebst einem Zwischenbe-

reich - wobei dieser als "Mitte" im intraphänomenalen Bereich gar nichts anderes bedeuten kann als Gleichabständigkeit zu den (wiederum phänomenalen) Polen (vgl. HELLER & KRÜGER 1974).
Daß die eigene Einschätzung der Differenzierungsfähigkeit bei eigenen Leistungen nicht unbedingt sehr aussagekräftig sein muß, ergab sich aus unseren Reaktionszeitversuchen: einige Vpn hatten das Gefühl, beim besten Willen nur Paarvergleiche anstellen, ihr Urteil immer nur auf den vorletzten Reiz beziehen zu können - die Urteilsverteilungen zeigten dennoch, daß ihren Leistungsbewertungen in Wirklichkeit sehr wohl ein absolutes System zugrunde gelegen haben muß.

6.3. Ausblick

Aus dem in den letzten Absätzen Gesagten werden gleichzeitig diejenigen Aspekte deutlich, denen unserer Meinung nach in erster Linie das weitere Augenmerk der Forschung auf dem zur Diskussion stehenden Untersuchungsgebiet gelten sollte. Der Deutlichkeit halber wollen wir sie hier noch einmal - jeweils für sich genommen - isoliert anführen:

(1) Zu nennen ist hier zunächst einmal der zuletzt behandelte und darum an dieser Stelle nur noch kurz anzusprechende Problemkreis der Anzahl der sinnvoll denkbaren Kategorialbereiche. Wir halten nach unseren Ergebnissen für wahrscheinlich, daß es hinsichtlich der Gütemaßstäbe für die eigene Leistung kognitiv nicht so viele Kategorialabschnitte, nicht so viele Differenzierungsmöglichkeiten gibt und geben kann wie bei der Beurteilung von Bleistiftlängen oder Gewichten etwa. Wir fragen daher: Stimmt dies? Gibt es tatsächlich mehr als drei kognitive Kategorialbereiche - wenn gleichzeitig das WITTEsche Theorem vom gleichen Umfang der Abschnitte noch Gültigkeit behalten soll?

(2) Relativ unproblematisch erscheint demgegenüber die Frage, ob wir mit unseren Ergebnissen, die die WITTEschen Annahmen zu bestätigen scheinen, wirklich Aussagen haben machen können oder ob es sich um zufällige Befunde handelt, die wir besonderen Umständen verdanken, etwa unseren speziellen Versuchsanordnungen oder unseren Versuchspersonen (fast alle waren schließlich Psychologie-Studenten älterer Semester). Wir fragen: Haben die WITTEschen Bezugssystem-Annahmen tatsächlich auch für den Bereich der Beurteilung der eigenen Leistung Gültigkeit oder sind unsere bestätigenden Ergebnisse, die sich in erster Linie auf Versuche mit niedrigen Kategorien-Anzahlen stützen, Zufallsprodukte?

(3) Abhängig sind die beiden genannten Fragen von der Wahl eines geeigneten Untersuchungs-"Materials". Dieses sollte nicht nur von sich aus hinreichend leistungsanregend und -fordernd sein, sondern darüber hinaus einen Weg aus dem Dilemma weisen, einerseits hinreichend viele Leistungsmöglichkeiten im Sinne von abgrenzbaren Klassen zu bieten, damit überhaupt ausdifferenzierte Urteile von den äußeren Gegebenheiten her möglich sind, so es sie inhaltlich gibt; andererseits aber soll dieses Material in seiner Mannigfaltigkeit auch mnestisch stabilisierbar sein - was wohl nur über begrenzte Anzahlen von Klassen geht. Wir sehen hier vorläufig keinen Weg und glauben dies weitgehend in der Natur des Untersuchungsgegenstandes "Leistungsverhalten" begründet. Daher hier unsere Frage nach der Möglichkeit eines Untersuchungsfeldes bzw. einer Versuchsanordnung, die diesem Problem abhilft.

(4) Bezüglich der Struktur der jeweiligen den Leistungsbeurteilungen zugrunde liegenden Gütemaßstäbe waren wir davon ausgegangen, daß sie sich in der Form von Differenzen

fassen lasse. Wenngleich sich dies - wie unsere Ergebnisse in ihrer Konsistenz dartun - insgesamt als tunlich und brauchbar erwies, so scheint uns dennoch nicht endgültig geklärt, ob und unter welchen Bedingungen der WEBER-Bruch, der ja immerhin in einer ganzen Reihe von Fällen seine Brauchbarkeit unter Beweis gestellt hat, heranzuziehen ist. Wir denken hier an die oben erörterten diesbezüglichen strukturellen Unterschiede in den Ereignismannigfaltigkeiten von Reaktionszeiten und Pfeilwurf-Punktwerten. Unsere Frage daher: Lassen sich weitere Ergebnisse finden, die das Verständnis der Sollwert-Istwert-Diskrepanzen bei der Beurteilung der eigenen Leistung als Differenz untermauern, oder sind letztlich auf diesem Gebiet andere Sichtweisen zweckmäßiger?

(5) Unsere letzte Frage gilt dem Problem der Partialsysteme auf dem Gebiete des Leistungsverhaltens. Wir hatten bei der Interpretation unserer Ergebnisse Grund zu der Annahme, daß bei der Leistungsbeurteilung Erfolgs- und Mißerfolgserleben nicht nur die zwei (bildlichen) Enden eines Kontinuums sind,sondern daß es - möglicherweise daneben - auch Bezugssysteme gibt, die nur den (kleineren oder größeren) Mißerfolg umfassen. Dies ergab sich daraus, daß wir nicht nur "zu große", sondern auch "zu kleine" Mißerfolge gefunden zu haben glauben. Möglicherweise gilt für den Fall des Erfolgserlebens analoges. Wir fragen also: Gibt es - eventuell neben einem Erfolgs-Mißerfolgs-Maßstab - getrennte Bezugssysteme für Erfolgs- und Mißerfolgsbeurteilung? Sind Erfolgs- und Mißerfolgserleben in dieser Hinsicht analog zu sehen oder gibt es - wie wir vermuten - strukturelle Unterschiede?

(6) Insgesamt scheint uns - gerade auf der Basis des zuletzt Angeführten - die Untersuchung des Mißerfolgsverhaltens

besonders fruchtbar zu sein. Wir glauben gesehen zu haben, daß Erfolg und Mißerfolg nicht direkt vergleichbar, sondern mindestens nach "leicht korrigierten" Maßstäben erlebt werden, daß das Wertungsgeschehen im Mißerfolgsfalle auf jeden Fall dynamischer ist - während der Erfolg eher als selbstverständlich hingenommen, relativ unreflektiert akzeptiert wird. Eine echte Auseinandersetzung mit der eigenen Leistung auf der Basis der relevanten Gütemaßstäbe findet dagegen primär bei Nicht-Erreichen der Ziele, beim Auftreten größerer Sollwert-Istwert-Diskrepanzen statt.

Betonenswert scheint uns hier abschließend vor allem noch eines: Wir haben im Verlauf dieser experimentellen Studie viel gelernt - nicht zuletzt, wie viele klangliche Nuancen es gibt bei Stoßseufzern wie "Oh nein!", "ach Gott!", "gerettet!" und was dergleichen spontane Äußerungen unserer Versuchspersonen zu einzelnen Leistungsergebnissen mehr waren - Belege in jedem Falle für die elementare Rolle der Auseinandersetzung mit der eigenen Leistung für unser Erleben und Verhalten, Belege gleichzeitig somit auch für die Bedeutung von Gütemaßstäben, von Bezugssystemstrukturen und -dynamiken bei der Beurteilung unserer Leistungen.

L I T E R A T U R - V E R Z E I C H N I S

Bochow, R. Sozialpsychologische Faktoren des Anspruchsniveaus. Psychol. Forsch., 1957, 25, 238-278

Forner, C. Prüfung und Prüfbarkeit von Hypothesen über mnestisch stabilisierte Bezugssysteme. Dissertation Münster 1974

Fuchs, R. Funktionsanalyse der Motivation. Z.exp.u.angew. Psychol., 1963, 10, 626-645

Heckhausen, H. Motivationsanalyse der Anspruchsniveau-Setzung. Psychol. Forsch., 1955, 25, 118-154

Heckhausen, H. Eine Rahmentheorie der Motivation in 10 Thesen. (1963a) Z.exp.u.angew. Psychol., 1963, 10, 604-626

Heckhausen, H. Hoffnung und Furcht in der Leistungsmotivation. (1963b) Meisenheim: Hain 1963

Heckhausen, H. Leistungsmotivation. in: Thomae, H. (Hg.) Handbuch der Psychologie, II. Göttingen: Hogrefe 1965, 602-702

Heckhausen, H. Allgemeine Psychologie in Experimenten. Göttingen: Hogrefe 1969

Heller, O. & H.-P. Krüger Direkte Skalierung in der Soziometrie. Psychol. Beitr., 1974, 16, 203-226

Hellpach, W. Sozialpsychologie. Stuttgart: Enke 1951

Helson, H. Adaptation level as a basis for a quantitative theory of frames of reference. Psychol. Rev., 1948, 55, 297-313

Hoppe, F. Erfolg und Mißerfolg. Psychol. Forsch., 1930, 14, 1-62

Kellner, B. Untersuchungen an bezugssystembildenden Mannigfaltigkeiten, die nicht auf physikalische Meßgrundlagen bezogen werden können. Zulassungsarbeit zur Diplom-Vorprüfung Münster 1966

Lewin, K. Feldtheorie in den Sozialwissenschaften. Herausgegeben von D. Cartwright. Stuttgart: Huber 1963

Martin, L.J. & G.E. Müller
Zur Analyse der Unterschiedsempfindlichkeit. Leipzig: Barth 1899

McClelland, D.G.; J.W. Atkinson; R.A. Clark & E.L. Lowell
The achievement motive. New York: Appleton 1953

Metzger, W. Psychologie. Darmstadt: Steinkopff 1975

Sievers, W. Leistungshöhe im Spiegel absoluter Urteile. Dissertation Münster 1970

Sitmalidis, S. Experimentelle und vergleichende Untersuchungen zum Voraussagewert von Helsons Adaptations- und Wittes Bezugssystemtheorie. Dissertation Münster 1974

Skroblin, B. Bezugssystembildung, Erfahrung und Motivation - am Beispiel von Lottogewinnen. Psychol. Beitr., 1972, 14, 376-407

Stevens, S.S. Adaptation level vs. the relativity of judgement. Amer. J. Psychol., 1958, 71, 633-646

Wever, E.G. & K.E. Zener
The method of absolute judgement in psychophysics. Psychol. Rev., 1928, 35, 466-493

Witte, W. Struktur, Dynamik und Genese von Bezugssystemen. (1960a) Psychol. Beitr., 1960, 4, 218-252

Witte, W. Über Phänomenskalen. (1960b) Psychol. Beitr., 1960, 4, 645-672

Witte, W. Das Problem der Bezugssysteme.
in: Metzger, W. & H. Erke (Hg.)
Handbuch der Psychologie, I,1.
Göttingen: Hogrefe 1966, 1003-1027

Witte, W. Zur Analyse der Absolutbeurteilungen sportlicher Leistungen.
Z.exp.u.angew. Psychol., 1971, 18, 678-691

Witte, W. Bezugssystem.
in: Ritter, J. (Hg.)
Historisches Wörterbuch der Philosophie, I.
Basel: Schwabe 1974, 910-912

Witte, W. Zum Gestalt- und Systemcharakter psychischer Bezugssysteme.
in: Ertel, S.; L. Kemmler & M. Stadler (Hg.)
Gestalttheorie in der modernen Psychologie.
Darmstadt: Steinkopff 1975, 76-93

Danksagung

Ich möchte an dieser Stelle nicht versäumen, meinen Versuchspersonen zu danken, die sich während der Durchführung der dieser Studie zugrunde liegenden Experimente recht eigentlich als Versuchspartner erwiesen haben.

Ganz besonderen Dank aber sage ich meinem verehrten Lehrer, Herrn Prof. Dr. W. Witte, aus dessen Forschungsarbeit heraus diese Untersuchung entstanden ist und dem ich eine Fülle wertvoller Anregungen verdanke, ohne die die Arbeit an diesem Thema kaum denkbar gewesen wäre.

Lebenslauf und Bildungsgang

Geboren am 26.4.44 in Herford als ältester von vier Söhnen des Bernard Brackhane und seiner Ehefrau Lotte, geb. Peter.

Zu Ostern 1950 Einschulung in die Katholische Volksschule Komturstraße in Herford, 1954 Wechsel auf das altsprachliche Friedrichs-Gymnasium in Herford, dort am 25.2.63 Abitur.
Von April 1963 bis März 1965 Militärzeit.
Seit dem Sommer-Semester 1965 Studium der Psychologie an der Westfälischen Wilhelms-Universität in Münster mit den Nebenfächern Philosophie, Pädagogik, Anatomie, Physiologie, Humangenetik und Psychiatrie. Diplom-Hauptprüfung am 13.2.70.

Während des Studiums seit dem Sommer-Semester 1966 studentische Hilfskraft in der Abteilung für Allgemeine Psychologie (Prof. Metzger), seit dem Sommer-Semester 1968 in der Abteilung für Angewandte Psychologie (Prof. Witte); dort auch nach bestandenem Examen wissenschaftliche Hilfskraft von März 1970 bis September 1972.
Von Oktober 1972 bis Oktober 1974 mit der Verwaltung der Dienstgeschäfte eines wissenschaftlichen Assistenten in der Abteilung für Allgemeine und Angewandte Psychologie (Prof. Witte) an der WWU Münster beauftragt, seit August 1973 zusätzlich mit der Wahrnehmung der Aufgaben des Kustos betraut.
Seit November 1974 Verwalter der Dienstgeschäfte eines wissenschaftlichen Assistenten am Lehrstuhl Psychologie II (Prof. Witte) der Universität Regensburg; dort auch im Winter-Semester 1975/76 wieder immatrikuliert für Psychologie mit den Nebenfächern Philosophie, Physiologie und Biologie.

A N H A N G

Protokolle der Experimente mit dem Dart-Spiel in der letzten Versuchsserie.

Anhang I:	Leistungsbeurteilungen nach 2 Kategorien
Anhang II:	Leistungsbeurteilungen nach 3 Kategorien
Anhang III:	Leistungsbeurteilungen nach 4 Kategorien
Anhang IV:	Leistungsbeurteilungen nach 5 Kategorien

Aufgeführt sind jeweils der Anspruchsniveau-Wert (AN)
der mit 6 Pfeilen erreichte Wert (Ist)
die Wertungen der Vpn (Wtg)

Anhang I 1

Versuchsperson LT, 1. Versuch, 2 Kategorien

\+ = gut

\- = schlecht

AN	Ist	Wtg	AN	Ist	Wtg
18	46	+		36	-
	45	+		33	-
30	45	+		40	-
35	21	-		21	-
	38	+		36	-
	34	-		34	-
	32	-		34	-
	46	+		27	-
	42	+		38	-
	39	+		18	-
40	37	-	35	46	+
	38	-		41	+
	37	-	40	42	+
	33	-		37	-
	41	+		31	-
	31	-		36	-
	41	+		44	+
	30	-		44	+
	25	-		40	+
	55	+		42	+
	32	-		40	+
	41	+	45	37	-
	41	+		50	+
	39	-		47	+
	48 *			41	-

*: das war mehr Zufall (Kommentar der Vp)

Anhang I 2

Versuchsperson RR, 1. Versuch, 2 Kategorien

\+ = gut
\- = schlecht

AN	Ist	Wtg
20	42	+
30	38	+
	32	+
35	24	-
	24	-
	31	+
	43	+
	37	+
	44	+
38	44	+
	45	+
	46	+
40	35	-
	23	-
	36	-
	50	+
	32	-
	47	+
42	38	-
	48	+
	35	-
	42	+
	43	+
	32	-
	46	+

AN	Ist	Wtg
	38	+
	44	+
	48	+
	48	+
45	49	+
	29	-
	40	+
	34	-
	43	+
	40	-
	47	+
	22	-
	25	-
	33	-
	31	-
	32	-
	44	+
	45	+
	33	-
	38	+
40	35	-
	33	-
	32	-
	44	+
	43	+

Anhang I 3

Versuchsperson GM, 1. Versuch, 2 Kategorien

\+ = gut
\- = schlecht

AN	Ist	Wtg	AN	Ist	Wtg
30	31	+		42	+
	40	+		18	-
35	28	-		17	-
	36	+		24	-
	30	-		20	-
	16	-		39	+
	40	+		38	+
	33	+		22	-
	24	-		31	-
	31	+		5	-
	27	-		25	-
	39	+		23	-
	44	+		31	-
38	33	-		36	+
	43	+		25	-
	37	+		27	-
	33	-		41	+
	16	-		14	-
	45	+		32	-
40	37	+		23	-
	15	-		34	+
	33	-		8	-
	41	+		29	-
	31	-		36	+
	32	-		44	+

Anhang I 4

Versuchsperson ND, 1. Versuch, 2 Kategorien

\+ = gut
\- = schlecht

AN	Ist	Wtg
40	33	+
	40	+
45	47	+
50	43	+
	43	+
	45	+
	30	-
	52	+
55	30	-
	30	-
	35	-
50	36	-
	43	+
	48	+
	45	+
	38	-
	24	-
	38	-
	47	+
	50	+
	49	+
	38	-
	47	+
	33	-
	47	+

AN	Ist	Wtg
	46	+
	34	-
	51	+
55	51	+
	5	-
	42	-
	31	-
	46	-
	34	-
	44	-
50	29	-
	39	-
	44	-
	40	-
	45	-
	43	-
	49	+
	45	-
	39	-
	40	-
	41	-
	39	-
	45	-
	52	+
	38	-

Anhang I 5

Versuchsperson JH, 1. Versuch, 2 Kategorien

\+ = gut
\- = schlecht

AN	Ist	Wtg
20-25	23	+
	28	+
30	28	+
	40	+
35	28	-
	34	+
	12	-
	30	+
	40	+
40	32	-
	22	-
	31	-
	46	+
45	26	-
	33	-
	34	+
	34	-
	21	-
	26	-
	37	+
	9	-
	36	+
	21	-
	34	-
	23	-

AN	Ist	Wtg
	33	+
	29	-
	43	+
	23	-
	3	+
	47	+
50	46	+
	24	-
	22	-
	41	+
	33	-
	40	+
	36	+
	37	+
	42	+
	45	+
	34	-
	29	-
	37	+
	30	-
	45	+
	45	+
	38	-
	46	+
	47	+

Anhang II 1

Versuchsperson LT, 2. Versuch, 3 Kategorien

+ = gut
0 = mittel
- = schlecht

AN	Ist	Wtg	AN	Ist	Wtg
35	31	-		37	-
	40	0		24	-
	36	0		29	-
	43	0		24	-
	19	-		38	0
	37	0		45	+
	40	0		34	-
40	37	0		45	
	42	0		33	-
	24	-		47	+
	30	-		38	0
	39	0		46	+
	30	-		37	0
	33	-		52	+
	28	-		47	+
	36	0		37	0
	42	+		39	0
	28	-		36	0
	41	0		38	0
	49	+		41	0
	36	0		44	+
	43	+		39	0
	40	+		34	-
	36	0		45	+
	36	0		40	0

Anhang II 2

Versuchsperson RR, 2. Versuch, 3 Kategorien

+ = gut
0 = mittel
- = schlecht

AN	Ist	Wtg
25	35	+
30	34	0
	35	0
33	33	0
	24	-
	21	-
	46	+
36	42	+
	8	-
	22	-
	19	-
	33	0
	42	+
39	29	0
	21	-
	28	0
	20	-
	30	0
	34	0
	47	+
	21	-
	34	+
	37	+
42	26	-
	25	-

AN	Ist	Wtg
	24	-
36	38	+
	25	-
	25	-
	23	-
	34	0
	36	+
	39	+
39	32	0
	26	-
	24	-
	13	-
	26	0
	25	-
	28	0
	42	+
	33	0
	17	-
	45	+
42	25	-
	32	0
	30	-
	28	-
	39	+
	38	+

Anhang II 3

Versuchsperson GM, 2. Versuch, 3 Kategorien

+ = gut
0 = mittel
- = schlecht

AN	Ist	Wtg
35	42	+
38	28	-
	46	+
	34	-
	46	+
40	32	-
	49	+
	30	-
	36	0
	39	0
	26	-
	47	+
42	41	0
	31	-
	29	-
	42	0
	45	+
	26	-
	41	0
	41	0
	33	-
	33	-
	43	+
	40	0
	36	0

AN	Ist	Wtg
	42	+
	25	-
	48	+
	32	-
	41	+
	31	-
	35	0
	39	0
	44	+
	41	+
	45	+
	35	-
	33	-
	43	+
	45	+
	38	0
	41	+
	36	0
	24	-
	27	-
	46	+
	48	+
44	39	0
	23	-
	40	0

Anhang II 4

Versuchsperson ND, 2. Versuch, 3 Kategorien

\+ = gut
0 = mittel
\- = schlecht

AN	Ist	Wtg
40	47	+
50	40	-
	51	+
52	32	-
50	40	-
	32	-
45	45	+
50	45	-
	40	-
	20	-
45	32	-
	34	-
40	35	-
	44	+
45	34	-
	32	-
	36	-
	50	+
50	39	-
	36	-
45	36	-
	31	-
	36	-
	35	-
40	40	+

AN	Ist	Wtg
45	23	-
	43	+
	37	-
	41	0
	35	-
	42	0
	46	
50	36	-
	51	
	32	-
	30	-
45	31	-
	30	-
	38	-
	50	+
50	37	-
	19	-
45	26	-
	34	-
40	36	0
	46	+
45	48	+
50	47	+
	31	-
	45	-

Anhang II 5

Versuchsperson JH, 2. Versuch, 3 Kategorien

+ = gut
0 = mittel
- = schlecht

AN	Ist	Wtg
30	36	+
40	46	+
50	39	0
	47	+
	38	0
	39	0
	48	+
	45	+
	42	0
	28	-
	36	-
	45	+
	32	-
	36	-
	38	-
	31	-
	34	-
	24	-
	42	+
	35	-
	34	-
	28	-
	38	-
	19	-
	40	0

AN	Ist	Wtg
	37	-
	33	-
	26	-
	28	-
	28	-
	27	-
	44	+
	43	+
	39	0
	28	-
	50	+
55	18	
	36	-
	44	+
	36	-
	40	-
	38	-
	31	-
	46	+
	26	-
	21	-
	43	+
	39	0
	38	-
	23	-

Anhang II 6

Versuchsperson JH, 4. Versuch, 3 Kategorien

+ = gut
0 = mittel
- = schlecht

AN	Ist	Wtg
50	43	+
	43	+
	18	-
	33	-
	53	+
55	35	-
	47	+
	27	-
	44	0
	36	-
	33	-
	50	+
	30	-
	45	0
	40	0
	42	0
	45	+
	43	0
	41	-
	21	-
	23	-
	50	+
	45	0
	36	-
	27	-

AN	Ist	Wtg
	35	-
	46	0
	33	-
	39	0
	32	-
	36	-
	39	0
	49	+
	31	-
	46	0
	41	0
	42	0
	49	+
	32	-
	34	-
	29	-
	30	-
	29	-
	43	0
	36	-
	20	-
	36	-
	23	-
	41	0
	37	-

Anhang III 1

Versuchsperson LT, 3. Versuch, 4 Kategorien

+ = gut
: = eher gut
. = eher schlecht
- = schlecht

AN	Ist	Wtg
40	57	+
	26	-
	33	.
	42	:
	41	:
	49	+
	27	-
	54	+
	43	:
	44	:
	33	.
	49	+
42	42	:
	38	-
	45	:
	30	-
	40	:
	49	+
	40	.
	32	-
	46	+
	42	:
	47	+
	37	.
	33	-

AN	Ist	Wtg
	41	:
	35	-
	42	:
	46	+
	39	.
	42	:
	54	+
	45	:
	47	+
45	41	:
	41	:
	43	:
	48	+
	37	.
	34	-
	44	:
	44	:
	49	+
	36	-
	40	:
	30	-
	49	+
	35	.
	30	-
	41	:

Anhang III 2

Versuchsperson RR, 3. Versuch, 4 Kategorien

\+ = gut
: = eher gut
. = eher schlecht
\- = schlecht

AN	Ist	Wtg
30	16	-
	24	.
	20	-
	18	-
	44	+
	33	+
	24	.
	19	-
	28	.
	34	:
	34	:
33	26	:
	28	:
	34	+
	25	.
	26	-
	31	:
	17	-
	36	+
	37	+
	30	:
	25	.
	31	:
	37	+
36	30	.

AN	Ist	Wtg
	39	+
	44	+
39	46	+
	34	.
	31	.
	51	+
	28	.
	39	:
	41	+
	37	:
	31	.
	28	-
	35	:
	33	.
42	30	.
	33	:
	42	+
	43	+
	38	:
	50	+
	33	.
	42	+
	30	-
	35	.
	38	:

Anhang III 3

Versuchsperson RR, 4. Versuch, 4 Kategorien

\+ = gut
: = eher gut
. = eher schlecht
\- = schlecht

AN	Ist	Wtg
35	42	+
	48	+
40	39	+
	30	:
	36	:
	38	:
	35	:
	43	+
	60	+
42	22	-
	43	+
	27	.
	42	+
	39	:
	34	.
	26	-
	39	:
	40	:
	39	:
	53	+
	33	-
	42	+
	36	-
	35	.
	27	-

AN	Ist	Wtg
	42	:
	45	+
45	40	:
	50	+
	38	.
	36	.
	39	:
	50	+
	38	:
	46	+
	42	.
	39	:
	39	.
	35	.
	39	:
	36	:
	34	-
	44	+
	44	:
	48	+
	43	:
	40	:
	46	+
48	39	-
	42	:

Anhang III 4

Versuchsperson GM, 3. Versuch, 4 Kategorien

+ = gut
: = eher gut
. = eher schlecht
- = schlecht

AN	Ist	Wtg
35	27	-
	39	:
	32	.
	12	-
	37	:
38	35	.
	33	-
	27	-
	44	+
	43	+
40	32	-
	26	-
	31	-
	32	-
	24	-
	41	+
	34	.
	37	.
	32	-
	29	-
	28	-
	38	:
	37	:
	33	.
	33	-

AN	Ist	Wtg
	24	-
	24	-
	25	-
	35	:
	36	:
	42	+
	21	-
	8	-
	29	.
	44	+
	43	+
	44	+
	27	-
	44	+
42	31	-
	18	-
	29	-
	46	+
	24	-
	43	+
	29	-
	34	.
	25	-
	20	-
	9	-

Anhang III 5

Versuchsperson GM, 4. Versuch, 4 Kategorien

+ = gut
: = eher gut
. = eher schlecht
- = schlecht

AN	Ist	Wtg	AN	Ist	Wtg
35	30	.		33	-
	25	-		39	+
	26	-		38	:
	34	:		24	-
	39	+		32	-
	27	-		25	-
	35	+		34	.
	42	+		28	-
38	26	-		14	-
	34	.		44	+
	38	+		19	-
	26	-		42	+
	35	:		39	:
	39	+		49	+
	28	-		37	:
	47	+		33	-
40	23	-		44	+
	28	-		39	:
	28	-		41	+
	27	-	42	30	-
	35	.		34	.
	44	+		34	.
	33	-		48	+
	28	-		40	:
	43	+		35	.

Anhang III 6

Versuchsperson ND, 3. Versuch, 4 Kategorien

+ = gut
: = eher gut
. = eher schlecht
- = schlecht

AN	Ist	Wtg	AN	Ist	Wtg	
40	38	+		35	.	
45	37	:		28	-	
	40	:		26	-	*
	31	.		42	+	
	30	-		33	.	
	43	+		19	-	
	34	.		25	-	
	34	.		16	-	
	33	.		40	:	
	28	-		24	-	
	45	+		34	.	
	26	-		22	-	
	17	-		22	-	
	12	-		34	:	
	28	-		47	+	
40	25	-		35	:	
	23	-		34	:	
	32	.		24	-	
	40	+		31	.	
45	47	+		40	+	
	36	:		34	.	
	43	+		37	:	
	40	+		37	:	
	51	+		14	-	
	37	:		40	+	

*: Vp wechselt die Pfeile

Anhang III 7

Versuchsperson JH, 3. Versuch, 4 Kategorien

\+ = gut
: = eher gut
. = eher schlecht
\- = schlecht

AN	Ist	Wtg	AN	Ist	Wtg
45	29	-		40	:
	33	:		40	:
	34	.		41	:
	46	+		45	+
50	31	-		47	+
	25	-		22	-
	17	-		47	+
	26	-		18	-
	26	-		26	-
	24	-		30	-
	29	-		24	-
	31	.		25	-
	43	:		30	.
	41	:		36	:
	21	-		44	+
	38	:		52	+
	36	.	55	37	:
	33	.		38	.
	34	.		41	:
	36	:		39	:
	29	-		27	-
	31	-		29	-
	39	:		37	:
	46	+		30	.
	32	-		36	.

Anhang IV

Versuchsperson LT, 4. Versuch, 5 Kategorien

++ = sehr gut
+ = gut
0 = mittel
- = schlecht
-- = sehr schlecht

AN	Ist	Wtg
40	42	+
	44	+
42	40	+
	44	++
	37	-
	19	--
	42	+
	38	-
	42	0
	51	++
	50	++
45	38	-
	38	--
	31	--
	55	++
	29	--
	36	-
	33	--
	31	--
	38	-
	46	+
	52	++
	52	++
	43	+
	47	+

AN	Ist	Wtg
	45	+
	42	0
	46	+
	46	+
	49	++
	44	0
	29	--
	40	0
	35	--
	51	++
	28	--
	38	-
	42	0
	35	--
	39	0
	45	+
	38	0
	30	--
	40	-
46	36	--
	50	++
	39	-
	48	+
	40	0
	43	0

SPERL